



# STIC Search Report

## Biotech-Chem Library

STIC Database Tracking Number: 111165

TO: Michael Borin  
Location: cm1/12a01/12d01  
Art Unit: 1631  
Monday, January 05, 2004

Case Serial Number: 09/930169

From: Toby Port  
Location: Biotech-Chem Library  
CM1-6A04  
Phone: 308-3534

[toby.port@uspto.gov](mailto:toby.port@uspto.gov)

### Search Notes

Dear Examiner Borin,

Here are the results of your search.  
Please feel free to contact me if you have any questions.

Toby Port

**THIS PAGE BLANK (USPTO)**

O'Bry n, Barbara

11/65

**From:** Borin, Michael  
**Sent:** Tuesday, December 30, 2003 3:14 PM  
**T :** STIC-Biotech/ChemLib  
**Subject:** Search request: 09/930169

Examiner: M.Borin  
CM1 12A01  
AU: 1631; Mailbox: 12D01

Tel.: 305-4506

RE: 09/930169; p43 fragment

---

Please conduct search of polypeptide SEQ ID 1,2 against the commercial and interference protein databases.

Thank you

**THIS PAGE BLANK (USPTO)**



	Region	/note= "beta6i strand"	75..77
FT	Region		79..85
FT	Region	/note= "beta7 strand"	
FT	Region	/note= "beta8 strand"	90..92
FT	Region	/note= "beta9 strand"	103..106
FT	Region	/note= "alpha 2 helix"	119..123
FT	Region	/note= "alpha 3 helix"	124..130
FT	Region	/note= "beta10 strand"	132..134
FT	Region	/note= "beta11 strand"	140..142
PX			
PX	(IMAG-) IMAGEGENE CO LTD.		
PI	Kim S., Ko Y., Kim YS., Jo YJ;		
XX	WPI; 2002-098017/13.		
DR	N-PADB; ABA34640.		
PT	Newel anti-tumour and anti-angiogenic agent of p43 comprises N-terminal domain and C-terminal domain containing eleven beta-strands forming a structural core and three flanking alpha-helices		
PS	Claim 1; Fig 1; 35bp; English.		
CC	The invention provides an anti-tumour and anti-angiogenic agent of p43 consisting of two domains, the N-terminal domain (146 amino acids) and C-terminal domain (166 amino acids) containing 11 beta-strands forming a structural core and 3 flanking alpha-helices. p43 is useful as an anti-tumour and anti-angiogenic agent. p43 and its C-terminal cytokine domain (EMAP II) induce regression of fibrosarcoma in immunocompromised mouse while its N-terminal domain does not. p43 is a potent cytokine as determined by the induction of tumour necrosis factor-alpha (TNF-alpha), interleukin-6 (IL-6), IL-8 and matrix metalloproteinase-9 or by its activity of chemotaxis. The present sequence represents the human p43 polypeptide.		
SC	Sequence	312 AA;	
DQ	Query Match	100.0%; Score 709; DB 23; Length 312; Best Local Similarity 100.0%; Pred. No. 4.3e-57; Matches 147; Conservative 0; Mismatches 0; Indels 0; Gaps 0	
OY	1 MANNDVLTRELKLEGGKAADQIEYLKKOVSILTKEKAIQAATLREKKLVENAKLKKEIE	60	
DB	1 MANNDAVTLRLBKKGKGAEDDIIETLYKKOVSLTKEKALIAQTIREKKLRVENAKLKKEIE	60	
OY	61 ELKOELIOAEIIONGVQKIAPFGSTPLIHANSWSENVISTAVTVSSGTKEQIKGTTGDE	120	
DB	61 ELLKOELLDAEIIONGVQKIAPFGSTGPLHANSWSENVISTAIVTVSSGTKEQIKGTTGDE	120	
OY	121 KKAKEKTEKKGEKKERKGOOSINGASDS	147	
DB	121 KKAKEKTEKKGEKKERKGOOSINGASDS	147	
RZ	AAB63244 standard; Protein; 328 AA.		

AC	AAB63244,
XX	
DT	26-MAR-2001 (first entry)
XX	
DE	Human breast cancer associated antigen protein sequence SEQ ID NO:606.
XX	
KM	Human; breast cancer; gastric cancer; prostate cancer; diagnosis;
KW	cancer associated antigen; cytostatic; cancer vaccine.
XX	
OS	Homo sapiens.
XX	
PN	WO200073801-A2.
XX	
XP	07-DEC-2000.
XX	
PF	26-MAY-2000; 2000MO-US14749.
XX	
PR	28-MAY-1999; 99US-0136526.
PR	10-SEP-1999; 99US-0153454.
PA	(LUDW-) LUDWIG INST CANCER RES.
XX	
PI	Obata Y;
XX	
DR	WPJ; 2001-025274/03.
XX	
PT	Nucleic acids encoding breast, gastric and prostate cancer associated
PT	antigen precursors, useful for diagnosing and treating a condition
PT	characterized by expression of an abnormal amount of a protein, e.g.
PT	cancer -
XX	
PS	Example 1; Page 478; 799pp; English.
XX	
CC	AAF22422 to AAF22626, AAF22627 to AAF22773 and AAF22774 to AAF23014
CC	represent nucleotide sequences encoding human breast, gastric and
CC	prostate cancer associated antigen precursors (CAAP) respectively.
CC	AAB63232 to AAB63467, AAB63468 to AAB63721 and AAB63722 to AAB63970
CC	respectively represent human breast, gastric and prostate CAAP protein sequence
CC	respectively. CAAPs have cytoskeletal activity and can be used in the
CC	production of cancer vaccines. The human CAAP proteins, peptides, nucleic
CC	acids or anti-CAAP antibodies are useful for diagnosing and treating a
CC	condition characterised by expression of an abnormal amount of a protein,
CC	e.g. cancer.
XX	
SQ	Sequence 328 AA;
XX	
Query Match	100.0%; Score 709; DB 22; Length 328;
Best Local Similarity	100.0%; Prid. No. 4,6e-57;
Matches 147; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
OY	1 MANNDATVLRLEQGKGAADQIIIEYLKQOVSLRKKALIQATLRKKLVENAKLKKEIE 60
DB	17 MANNDAIVLRLEQGGAGADQIIEVLKQOVSLRKKALIQATLRKKLVENAKLKKEIE 76
OY	61 ELKQELIQAEILONGVKQIAFPSSGTPHLANSWSVENVIQSTAVTVSSGTKEQIKGTTGE 120
DB	77 ELKQELIQAEILONGVKQIAFPSSGTPHLANSWSVENVIQSTAVTVSSGTKEQIKGTTGE 136
OY	121 KKAKEIKKK 147
DB	137 KKAKEIKKK 163
RESULT 3	
ID	AA72577
XX	AA72577 standard; Protein: 310 AA.
AC	AA72577;
DT	25-MAR-2003 (updated)
DT	29-SEP-1995 (first entry)
XX	
DE	Mouse EMAPIT.

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2004, 18:17:12 ; Search time 11.5294 Seconds  
(without alignments)  
599.590 Million cell updates/sec

Title: US-09-930-169-1  
Perfect score: 709  
Sequence: 1 MANNDAVLKRLKQKGAEDQ.....EKGKKEKQKQSIAGSADS 147

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 127863 seqs, 47026705 residues  
Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_41.\*  
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	709	100.0	312 1	MCAL HUMAN
2	548.5	77.4	539 1	MCAL CRICK
3	509	71.8	310 1	MCAL MOUSE
4	125.5	17.7	516 1	P54 ENTFC
5	117	16.5	662 1	TLPB BACSU
6	116.5	16.4	1079 1	IF2P SCHRO
7	108.5	15.3	662 1	MCPE BACSU
8	106.5	15.0	1093 1	TWFI HUMAN
9	106	15.0	461 1	US45 LACLC
10	104	14.7	473 1	YVCE BACSU
11	104	14.7	570 1	STIM DROME
12	104	14.7	961 1	VDP BOVIN
13	103.5	14.6	561 1	EXBA BACGD
14	103	14.5	1130 1	Y117 CAEBL
15	103	14.5	1169 1	SMC METUA
16	103	14.5	2230 1	GOG4 HUMAN
17	102	14.4	886 1	RA50 SUTAC
18	102	14.4	1220 1	IF2P HUMAN
19	101.5	14.3	539 1	MS33 HYDAT
20	101.5	14.3	1068 1	H1PR HUMAN
21	101	14.2	956 1	KF5C MOUSE
22	101	14.2	957 1	KF5C HUMAN
23	101	14.2	1938 1	MYSS CHICK
24	100.5	14.2	1690 1	C190 DROME
25	100	14.1	1243 1	SMC4 MICRA
26	99.5	14.0	886 1	RA50 ARCFU
27	99.5	14.0	2116 1	MY52 DICDI
28	99	14.0	1940 1	MYH3 CHICK
29	98.5	13.9	407 1	Y173 AQUAE
30	98.5	13.9	879 1	RA50 SUTLO
31	98	13.8	919 1	INCE HUMAN
32	98	13.8	1197 1	YOIK ECOLI
33	97.5	13.8	553 1	YOIK ECOLI

34	97.5	13.8	727 1	MEP1 ARATH	Q91w65 arabidopsis
35	97	13.7	1605 1	RBBI MOUSE	Q99p15 mus musculus
36	97	13.7	2434 1	YCF1 OENHO	Q9mth5 oenothera h
37	96.5	13.6	843 1	YMS1 YEAST	Q05050 saccharomyc
38	96	13.5	856 1	CLPB HELPY	P71404 helicobacte
39	96	13.5	1526 1	MY52 SCHRO	Q9u516 schistosach
40	96	13.5	1534 1	RRBI CANFA	Q28298 canis famli
41	95.5	13.5	793 1	CALD HUMAN	Q05682 homo sapien
42	95.5	13.5	1084 1	MYSS RABIT	P02562 oryctolagus
43	95.5	13.5	1447 1	SGS1 YEAST	P35187 saccharomyc
44	95.5	13.5	3214 1	BPAL HUMAN	Q03001 homo sapien
45	95	13.4	579 1	G160 HUMAN	Q08378 homo sapien

ALIGNMENTS

RESULT 1  
MCAL\_HUMAN STANDARD; PRT; 312 AA.  
ID MCAL\_HUMAN  
AC Q12904; Q96C09; 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Multienzyme complex auxiliary component p43 [Contains:  
DE Endothelial-monocyte activating polypeptide II (EMAP-II) (small  
DE inducible cytokine subfamily B member 1)].  
GN SCF1 OR EMAP2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxId=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95014290; PubMed=7929199;  
RA Kao J., Houck K., Fan Y., Haehnel I., Libutti S.K., Kayton M.L.,  
RA Grikscheit T., Chabot J., Nowygrod R., Greenberg S., Khang W.J.,  
RA Leung D.W., Hayward J.R., Kiesel W., Heath M., Brett J., Stern D.M.;  
RT "Characterization of a novel tumor-derived cytokine. Endothelial-  
RT monocyte activating polypeptide II,"  
RL J. Biol. Chem. 269:25106-25119(1994)  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Hsieh F.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marinsina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stjepeticen M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,  
RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Pringle C.,  
RA Rana S.S., Luquellano N.A., Peters G.J., Abramson R.D., Mulvaney S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Wooley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Kesteman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzyzanski M.I., Skalek U., Smalins D.E.,  
RA Schnerch A., Schein J.E., Jones S.U.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences,"  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
CC -I- SIMILARITY: Contains 1 tRNA-binding domain.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

**THIS PAGE BLANK (USPTO)**

or send an email to [license@ib-sib.ch](mailto:license@ib-sib.ch)).

CC or send an email to [license@ib-sib.ch](mailto:license@ib-sib.ch)).

CC EMBL; U10117; AAA62202.1; -

DR EMBL; BC014051; AAA14051.1; -

DR PDB; 1E7Z; 06-FEB-01.

DR PDB; 1EU7; 06-SEP-00.

DR PDB; 1FL0; 07-FEB-01.

DR Genew; HGNC:10648; SCYE1.

DR MIM; 603605; -

DR GO; GO:0005615; C:extracellular space; TAS.

DR GO; GO:0005125; F:cytokine activity; TAS.

DR GO; GO:0000049; F:RNA binding activity; TAS.

DR GO; GO:0006418; P:amino acid activation; TAS.

DR GO; GO:0006935; P:chemotaxis; TAS.

DR GO; GO:0006954; P:inflammatory response; TAS.

DR InterPro; IPR002547; RNA bind.

DR Pfam; PF01588; RNA bind.

DR PIRSF; PIRSF005381; EMAP11; 1.

DR PROSITE; PSS0886; TRBD; 1.

DR Protein biosynthesis; RNA-binding; Cytokine;

KM Protein biosynthesis; RNA-binding; Cytokine;

KM 3D-structure.

FT PROPEP 1 146

FT CHAIN 147 312

FT DOMAIN 151 252

FT CONFLICT 79 79

FT SEQUENCE 312 AA; 34326 MW; 946310A0216F7587 CRC64;

Query Match Best Local Similarity 100.0%; Score 709; DB 1; Length 312; Matches 147; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MANDAVLKRLEBKGAEADQIIIEYLRQVSLKKEKALIQATLRREKKLRVENAKLKEIE 60

DB 1 MANDAVLKRLEBKGAEADQIIIEYLRQVSLKKEKALIQATLRREKKLRVENAKLKEIE 60

QY 61 ELKQELIOAEIONGVQKQIAPFSGTPLHANSVSNVIOSTAVTTVSSGTKEQIKGTGDE 120

DB 61 ELKQELIOAEIONGVQKQIAPFSGTPLHANSVSNVIOSTAVTTVSSGTKEQIKGTGDE 120

QY 121 KKAKEKIEKGEKKEKKQKQSIAGSADS 147

DB 121 KKAKEKIEKGEKKEKKQKQSIAGSADS 147

## RESULT 2

MCAL CRIGR STANDARD; PRT; 359 AA.

AC 054873;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Multisynthetase complex auxiliary component p43 [contains: Endothelial-monocyte activating polypeptide II (EMAP-II) (Small inducible cytokine subfamily E member 1)].

GN SCYE1.

OS Cricetulus griseus (Chinese hamster).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Cricetulus.

OC NCBI\_TaxID=10029;

QX [1]

RN SEQUENCE FROM N.A.

RP MEDLINE=98070438; PubMed=9405472;

RA Quevillon S., Agou F., Robinson J.-C., Mirande M.;

RT "The p43 component of the mammalian multi-synthetase complex is likely to be the precursor of the endothelial monocyte-activating polypeptide II cytokine.";

RL J. Biol. Chem. 272:32573-32579(1997).

CC -1- SUBUNIT: COMPONENT OF THE MULTISYNTHETASE COMPLEX WHICH IS COMPRISED OF A BIFUNCTIONAL GLUTAMYL-PROLYL-TRNA SYNTHETASE, THE MONOSPECIFIC ISOENZYME, LEUCYL, GLUTAMINYL, METHIONYL, LYSYL, ARGINYL, AND ASPARYL-TRNA SYNTHETASES AS WELL AS THREE AUXILIARY

CC PROTEINS, P18, P48 AND P43.

CC -1- SIMILARITY: Contains 1 tRNA-binding domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@ib-sib.ch](mailto:license@ib-sib.ch)).

CC EMBL; AF021800; AA95207.1; -

DR InterPro; IPR002547; RNA bind.

DR Pfam; PF01588; RNA bind.

DR PIRSF; PIRSF005381; EMAP11; 1.

DR PROSITE; PSS0886; TRBD; 1.

KM Protein biosynthesis; RNA-binding; Cytokine.

FT DOMAIN 198 299

FT SEQUENCE 359 AA; 39601 MW; 4D868D1B65D72C23 CRC64;

Query Match Best Local Similarity 77.4%; Score 548.5; DB 1; Length 359; Matches 119; Conservative 10; Mismatches 17; Indels 3; Gaps 2;

QY 1 MANDAVLKRLEBKGAEADQIIIEYLRQVSLKKEKALIQATLRREKKLRVENAKLKEIE 60

DB 47 MANDAVLKRLEBKGAEADQIIIEYLRQVSLKKEKALIQATLRREKKLRVENAKLKEIE 106

QY 61 ELKQELIOAEIONGVQKQIAPFSGTPLHANSVSNVIOSTAVTTVSSGTKEQIKGTG 118

DB 107 ELKQELIOAEIONGVQKQIAPFSGTPLHANSVSNVIOSTAVTTVSSGTKEQIKGTG 165

QY 119 DEKKAKEKIEKGEKKEKKQKQSIAGSADS 147

DB 166 DEKKAKEKIEKGEKKEKKQKQSIAGSADS 194

## RESULT 3

MCAL MOUSE STANDARD; PRT; 310 AA.

AC P31230; Q06659;

DT 01-JUL-1993 (Rel. 26, Created)

DT 01-JUL-1993 (Rel. 26, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Multisynthetase complex auxiliary component p43 [contains: Endothelial-monocyte activating polypeptide II (EMAP-II) (Small inducible cytokine subfamily E member 1)].

GN SCYE1 OR EMAP2.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OC NCBI\_TaxID=10090;

QX [1]

RN SEQUENCE FROM N.A.

RP MEDLINE=95014290; PubMed=7929199;

RA Kao J., Houck K., Fan Y., Haehnle I., Libutti S.K., Kayton M.L.,

RA Griskevicius T., Chabot J., Nowygrad R., Greenberg S., Xiang W.J.,

RA Leung D.W., Hayward J.R., Kistner W., Heath M., Bretz J., Stern D.M.,

RT "Characterization of a novel tumor-derived cytokine, Endothelial-monocyte activating polypeptide II.";

RL J. Biol. Chem. 269:25106-25119(1994).

RN [2]

RP SEQUENCE FROM N.A.

RP MEDLINE=22388257; PubMed=12477932;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

RA Stapleton M., Soares M.B., Bonaldo M.F., Casavani T.L., Scheetz T.E.,

RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,

RA Rana S.S., Loggiano N.A., Peters G.J., Abramson R.D., Millay S.J.,

RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

**THIS PAGE BLANK (USPTO)**



```

FT Region 75..77 /note= "beta1 strand"
FT /note= "beta7 strand"
FT Region 79..85 /note= "beta7 strand"
FT Region 90..92 /note= "beta8 strand"
FT Region 103..106 /note= "beta9 strand"
FT Region 119..123 /note= "alpha 2 helix"
FT Region 124..130 /note= "alpha 3 helix"
FT Region 132..134 /note= "beta10 strand"
FT Region 140..142 /note= "beta11 strand"
FT Region 140..142 /note= "beta11 strand"
FT WO200195927-A1.
PN 20-DEC-2001.
PD 14-JUN-2000; 2000WO-KR00630.
XX 14-JUN-2000; 2000WO-KR00630.
XX 14-JUN-2000; 2000WO-KR00630.
XX (IMAG-) IMAGENE CO LTD.
XX Kim S, Ko Y, Kim YS, Jo YJ,
PI WPI; 2002-098017/13.
XX N-PSDB; ABA94640.
XX
XX Novel anti-tumour and anti-angiogenic agent of p43 comprises N-terminal
PT domain and C-terminal domain containing eleven beta-strands forming a
PT structural core and three flanking alpha-helices
XX
XX Claim 1; Fig 1; 35pp; English.
XX
XX The invention provides an anti-tumour and anti-angiogenic agent of p43
XX consisting of two domains, the N-terminal domain (146 amino acids) and
XX C-terminal domain (166 amino acids) containing 11 beta-strands forming
XX a structural core and 3 flanking alpha-helices. p43 is useful as an anti-
XX -tumour and anti-angiogenic agent. p43 and its C-terminal cytokine domain
XX (EMAP II) induce regression of fibrosarcoma in immunocompromised mouse
XX while its N-terminal domain does not. p43 is a potent cytokine as
XX determined by the induction of tumour necrosis factor-alpha (TNF-alpha),
XX interleukin-6 (IL-6), IL-8 and matrix metalloproteinase-9 or by its
XX activity of chemotaxis. The present sequence represents the human p43
XX polypeptide.
XX
XX Sequence 312 AA;
SQ
Query Match 100.0%; Score 709; DB 23; Length 312;
Best Local Similarity 100.0%; Pred. No. 4.3e-57;
Matches 147; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MANNDAVLRLEKKGAEADQIIEYLKQVSLLEKKAIIQATLREKKLRVENAKLKEIE 60
DB 1 MANNDAVLRLEKKGAEADQIIEYLKQVSLLEKKAIIQATLREKKLRVENAKLKEIE 60
QY 61 ELKQELIQAEIQNGVQKQIAFPSTGPLHANSMSENVISQTAVTTVSSGTKEQIKGTTGDE 120
DB 61 ELKQELIQAEIQNGVQKQIAFPSTGPLHANSMSENVISQTAVTTVSSGTKEQIKGTTGDE 120
QY 121 KKAKEKIEKKGEKKKKQKQSIAGSADS 147
DB 121 KKAKEKIEKKGEKKKKQKQSIAGSADS 147
XX
XX RESULT 2
XX AAB63244
XX ID AAB63244 standard; Protein; 328 AA.
XX

```

```

AC AAB63244;
XX
XX 26-MAR-2001 (first entry)
XX
XX Human breast cancer associated antigen protein sequence SEQ ID NO:606.
XX
XX Human; breast cancer; gastric cancer; prostate cancer; diagnosis;
XX cancer associated antigen; cytostatic; cancer vaccine.
XX
XX Homo sapiens.
XX
XX WO200073801-A2.
XX
XX 07-DEC-2000.
XX
XX 26-MAY-2000; 2000WO-US14749.
XX
XX 28-MAY-1999; 99US-0136526.
XX 10-SEP-1999; 99US-0153454.
XX
XX (LUDWIG INST CANCER RES.
XX
XX Odata Y;
XX
XX WPI; 2001-025274/03.
XX
XX Nucleic acids encoding breast, gastric and prostate cancer associated
XX antigen precursors, useful for diagnosing and treating a condition
XX characterized by expression of an abnormal amount of a protein, e.g.
XX cancer -
XX
XX Example 1; Page 478; 799pp; English.
XX
XX AAF22422 to AAF22626, AAF22627 to AAF22773 and AAF22774 to AAF23014
XX represent nucleotide sequences encoding human breast, gastric and
XX prostate cancer associated antigen precursors (CAAP) respectively.
XX AAB63232 to AAB63467, AAB63468 to AAB63721 and AAB63722 to AAB63970
XX represent human breast, gastric and prostate CAAP protein sequence
XX respectively. CAAPs have cytostatic activity and can be used in the
XX production of cancer vaccines. The human CAAP proteins, peptides, nucleic
XX acids or anti-CAAP antibodies are useful for diagnosing and treating a
XX condition characterized by expression of an abnormal amount of a protein,
XX e.g. cancer.
XX
XX Sequence 328 AA;
SQ
Query Match 100.0%; Score 709; DB 22; Length 328;
Best Local Similarity 100.0%; Pred. No. 4.6e-57;
Matches 147; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MANNDAVLRLEKKGAEADQIIEYLKQVSLLEKKAIIQATLREKKLRVENAKLKEIE 60
DB 17 MANNDAVLRLEKKGAEADQIIEYLKQVSLLEKKAIIQATLREKKLRVENAKLKEIE 76
QY 61 ELKQELIQAEIQNGVQKQIAFPSTGPLHANSMSENVISQTAVTTVSSGTKEQIKGTTGDE 120
DB 77 ELKQELIQAEIQNGVQKQIAFPSTGPLHANSMSENVISQTAVTTVSSGTKEQIKGTTGDE 136
QY 121 KKAKEKIEKKGEKKKKQKQSIAGSADS 147
DB 137 KKAKEKIEKKGEKKKKQKQSIAGSADS 163
XX
XX RESULT 3
XX AAR72577
XX ID AAR72577 standard; Protein; 310 AA.
XX
XX AAR72577;
XX
XX 25-MAR-2003 (updated)
XX 29-SEP-1995 (first entry)
XX
XX Mouse EMAP11.
XX

```



```

XX EMAP11; endothelial monocyte activating polypeptide II; chemotaxis;
KW inflammation; tissue factor; tumor; cancer; therapy; metha;
KM sarcoma.
XX
OS Mus musculus.
XX
PN M09509180-A1.
XX
PD 06-APR-1995.
XX
PF 29-SEP-1994; 94MO-US11085.
XX
PR 29-SEP-1993; 93US-0129456.
XX
PA (UYCO ) UNIV COLUMBIA NEW YORK.
XX
PI Claus M, Kao J, Kayton M, Libutti SK, Stern DM;
XX
DR WPI; 1995-147389/19.
DR N-PSDB; AAQ86718.
XX
PT New endothelial monocyte activating polypeptide II - induces
PT chemotaxis, inflammation and tissue factor, useful for treating
PT tumours, also related antibodies, DNA and active fragments
XX
PS Disclosure; Fig.4; 180pp; English.
XX
XX A mouse metha sarcoma CDNA library was screened with a probe
CC based on the N-terminal sequence of mouse EMAP11. Overlapping
CC clones were combined to obtain a contiguous full-length sequence
CC (given in AAQ86718) encoding a 33 kDa protein (AAR72577).
CC Recombinant EMAP11 was expressed in E. coli.
CC (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ Sequence 310 AA;
Query Match 71.8%; Score 509; DB 16; Length 310;
Best Local Similarity 76.4%; Pred. No. 1,1e-38;
Matches 113; Conservative 8; Mismatches 23; Indels 4; Gaps 3;
QY 1 MANDAVLKRLQKGAENDQIIIEYKQGVSLKEKALIQATLREKKLRVENAKLKEIE 60
DB 1 MATNDVAVLKRLQKGAENDQIIIEYKQGVSLKEKALIQATLREKKLRVENAKLKEIE 60
QY 61 ELKQELIQAETONGVKQIAFPSPGPIHANSWSENVIOSTAV-TTVSSGTQKQIKGTGD 119
DB 61 ELKQELIQAETONGVKQIAFPSPGPIHANSWSENVIOSTAV-TTVSSGTQKQIKGTGD 119
QY 120 EKKAKKELKKKKKKKKKKQKSIAGSADS 147
DB 119 EKKYKEKTEKKKKKKKKKKQKSIAGSADS 145
RESULT 4
ABBS9210
ID ABBS9210 standard; Protein; 294 AA.
XX
AC ABBS9210;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 4422.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN M0200171042-A2.
XX
PD 27-SEP-2001.

```

```

PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL03313.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 4422; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutic and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABBS7737-ABBS72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 294 AA;
Query Match 16.9%; Score 119.5; DB 22; Length 294;
Best Local Similarity 24.8%; Pred. No. 0.007;
Matches 37; Conservative 33; Mismatches 52; Indels 27; Gaps 4;
QY 6 AVLRLEQKGAENDQIIIEYKQGVSLKEKALIQATLREKKLRVENAKLKEIEELKOE 65
DB 2 ADLQGISNNRERAILNSTIEALISGIGQ-----QLVERQKQELKENNALAKEVYALAQ 57
QY 66 LIQAEIONGVKQIAFP-----SGTPIHANSWSENVIOSTAVTVSSGTQKQIKGTG 118
DB 58 LVQAELENGKQKQIPVPGARFCTSAAPV---VMPARAGPATAPAPA----- 102
QY 119 DEKAKKELKKKKKKKKKKQKSIAGSADS 147
DB 103 -PKPAKPKKKKKKKKKPAKPAAPEA 130
RESULT 5
AAM78507
ID AAM78507 standard; Protein; 1047 AA.
XX
AC AAM78507;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human protein SEQ ID NO 1169.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; hematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukemia;
KW nervous system disorder; arthritis; inflammation.
XX
OS Homo sapiens.
XX
PN M0200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US04098.
XX
PR 03-FEB-2000; 2000US-0496914.

```

PR 27-APR-2000; 2000US-0560875.  
 PR 20-JUN-2000; 2000US-0598075.  
 PR 19-JUL-2000; 2000US-0620325.  
 PR 01-SEP-2000; 2000US-0654936.  
 PR 15-SEP-2000; 2000US-0663561.  
 PR 20-OCT-2000; 2000US-0693325.  
 PR 30-NOV-2000; 2000US-0728422.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;  
 PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
 PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;  
 XX WPI; 2001-476283/51.  
 DR N-PSDB; AAK51640.

PT Nucleic acids encoding polypeptides with cytokine-like activities,  
 PT useful in diagnosis and gene therapy -  
 XX

PS Claim 20; Page 3407-3409; 6221pp; English.

CC The invention relates to polynucleotides (AAK51456-AAK53435) and the  
 CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activity/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukemia, nervous system disorders, arthritis and  
 CC inflammation.  
 CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
 CC (AAM80020) are omitted as the relevant pages from the sequence listing  
 CC were missing at the time of publication.  
 CC

XX Sequence 1047 AA;

Query Match 15.7%; Score 111.5; DB 22; Length 1047;

Best Local Similarity 23.0%; Pred. No. 0.2;

Matches 38; Conservative 43; Mismatches 49; Indels 35; Gaps 5;

QY 8 IRLRLEOKGAEADQIIIEYLKQOVSLKEKAILQATLREKKLRVNAKLEIELEKQELI 67

DB 736 LKLDVYRQQAQAIIFLKEQISLAERKM-----DYRLQRAEAGQGEVSLREKLL 769

QY 68 QAEIQ-NGVQKQIAFPSTPLHANSWSENVIOS-----TAVTVSSG 108

DB 790 VAENRLQAVFALCSSQHTHMIENSDISEFTIRKETEVEGLQDKLNKRDKEVTALTSQTEM 849

QY 109 TKEQI-----KGTGDEK-----KAKEKIEKKGEKKKKQOOSIAG 143

DB 850 LRAQVSALESCKSGEKVKDALKEKRRLAELETVSRKTHDASG 894

RESULT 6  
 AAM79491

ID AAM79491 standard; Protein; 1047 AA.

XX AAM79491;

DT 06-NOV-2001 (first entry)

XX Human protein SEQ ID NO 3137.

XX Human, cytokine; cell proliferation; cell differentiation; gene therapy;  
 KM vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KM tissue growth factor; immunomodulatory; cancer; leukemia;  
 KM nervous system disorder; arthritis; inflammation.

XX Homo sapiens.

PN W0200157190-A2.

XX 09-AUG-2001.

PF 05-FEB-2001; 2001WO-US04098.

XX 03-FEB-2000; 2000US-0496914.

PR 27-APR-2000; 2000US-0560875.

PR 20-JUN-2000; 2000US-0598075.

PR 19-JUL-2000; 2000US-0620325.

PR 01-SEP-2000; 2000US-0654936.

PR 15-SEP-2000; 2000US-0663561.

PR 20-OCT-2000; 2000US-0693325.

PR 30-NOV-2000; 2000US-0728422.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;  
 PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
 PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;  
 XX WPI; 2001-476283/51.  
 DR N-PSDB; AAK52624.

DR Claim 20; Page 262; 6221pp; English.

CC The invention relates to polynucleotides (AAK51456-AAK53435) and the  
 CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activity/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukemia, nervous system disorders, arthritis and  
 CC inflammation.  
 CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
 CC (AAM80020) are omitted as the relevant pages from the sequence listing  
 CC were missing at the time of publication.  
 CC

XX Sequence 1047 AA;

Query Match 15.7%; Score 111.5; DB 22; Length 1047;

Best Local Similarity 23.0%; Pred. No. 0.2;

Matches 38; Conservative 43; Mismatches 49; Indels 35; Gaps 5;

QY 8 IRLRLEOKGAEADQIIIEYLKQOVSLKEKAILQATLREKKLRVNAKLEIELEKQELI 67

DB 736 LKLDVYRQQAQAIIFLKEQISLAERKM-----DYRLQRAEAGQGEVSLREKLL 769

QY 68 QAEIQ-NGVQKQIAFPSTPLHANSWSENVIOS-----TAVTVSSG 108

DB 790 VAENRLQAVFALCSSQHTHMIENSDISEFTIRKETEVEGLQDKLNKRDKEVTALTSQTEM 849

QY 109 TKEQI-----KGTGDEK-----KAKEKIEKKGEKKKKQOOSIAG 143

DB 850 LRAQVSALESCKSGEKVKDALKEKRRLAELETVSRKTHDASG 894

RESULT 7  
 AAR14150

ID AAR14150 standard; Protein; 461 AA.

XX AAR14150;

DT 25-MAR-2003 (updated)

DT 09-DEC-1991 (first entry)

XX MSP encoded by pUCRS (DSM 5803).

[illegible]

PN NL9000753-A.  
 XX 16-OCT-1991.  
 XX  
 PF 30-MAR-1990; 90NL-0000753.  
 XX  
 PR 30-MAR-1990; 90NL-0000753.  
 XX  
 PA (NEZU-) NEDERLANDS INST ZUI.  
 XX  
 DR WPI; 1991-330695/45.  
 XX  
 DR N-PSDB; AAQ14501.  
 XX  
 PT Lactococcus lactis DNA fragments - contg. extracellular protein  
 PT signal peptide sequence  
 XX  
 PS Disclosure; Fig 4; 20pp; Dutch.  
 XX  
 CC The protein has a mol.wt. of 60 kD and is encoded by the KpnI/ClaI  
 CC fragment from pNZ1011. Lactococci transformed with the usp45 gene may  
 CC be used to produce extracellular heterologous enzymes of use in the food  
 CC industry, e.g. chymosin or alpha-amylase.  
 CC See also AAQ14502.  
 CC (Updated on 25-MAR-2003 to correct PF field.)  
 CC  
 SQ Sequence 461 AA;  
  
 Query Match 15.2%; Score 108; DB 12; Length 461;  
 Best Local Similarity 26.5%; Pred. No. 0.15; Mismatches 69; Indels 12; Gaps 5;  
 Matches 41; Conservative 33;  
  
 QY 1 MANNDVAVLKRLKQGAADQIIEYLKQOVSILKE-----KAIQATLREKKLRVENAKL 55  
 DB 33 IAKQDAPISAGAKAQAOQVDSLOSKVDSLOQKQSTYAKIASEKAKALNAQATL 92  
 QY 56 KKEIEIKQELIOAEIQNGVKQIAFPSTPLHANSVSENVIS-TAVTVSSGTKEQIK 114  
 DB 93 NESIKK-RTKTLBAQARSQVNSATNYMDAVVNSKSLTDVIGKVTATVSSANKQMLE 151  
 QY 115 GGTGDEKAKKEIEKGEKKEKQO--SINGSADS 147  
 DB 152 ---QOEKQKELSQKSETVKKNYNQVLSQSLDS 183  
  
 RESULT 10  
 AAR42818  
 ID AAR42818 standard; Protein; 1093 AA.  
 XX  
 AC AAR42818;  
 XX  
 DT 25-MAR-2003 (updated)  
 DT 27-APR-1994 (first entry)  
 XX  
 DE TMF.  
 XX  
 KW TATA modulating factor; TMF; transcription; TATA box; promoter; HIV-1;  
 KW human immunodeficiency virus-1; short arm; human chromosome 3; p12-p21;  
 KW translocation; cancer.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Region 437..850  
 FT /label= TATA binding region  
 FT Region 769..777  
 FT /note= "Ubiquitin-mediated protein degradation  
 FT consensus sequence homology region"  
 FT 454..614  
 FT /note= "Region with Leucine zipper secondary  
 FT structure"  
 FT Region 986..1069  
 FT /note= "Region with Leucine zipper secondary  
 FT structure"

FT Region 1070..1078  
 FT /note= "Ubiquitin-mediated protein degradation  
 FT consensus sequence homology region"  
 XX  
 PN MO9320106-A1.  
 XX  
 PD 14-OCT-1993.  
 XX  
 PF 31-MAR-1993; 93MO-US03077.  
 XX  
 PR 02-APR-1992; 92US-0862025.  
 XX  
 PA (TEXA) UNIV TEXAS SYSTEM.  
 XX  
 PI Gaynor RB, Wu F;  
 XX  
 DR WPI; 1993-336836/42.  
 DR N-PSDB; AAQ49397.  
 XX  
 PT New protein cellular factor - capable of binding double stranded  
 PT HIV-1 tata region and activating gene expression of HIV-LTR  
 XX  
 PS Claim 2; Fig 1; 75pp; English.  
 XX  
 CC This sequence represents TATA modulating factor (TMF). TMF is a  
 CC protein of mol. wt. 123-130 kD which activates transcription in most  
 CC genes, esp. in human immunodeficiency virus-1 (HIV-1) by binding to  
 CC the TATA box region of the promoter. TMF is encoded by the short  
 CC arm of human chromosome 3 in the region p12-p21 which is often  
 CC involved in translocations in patients having lung and other types  
 CC of cancer.  
 CC (Updated on 25-MAR-2003 to correct PN field.)  
 CC  
 SQ Sequence 1093 AA;  
  
 Query Match 15.0%; Score 106.5; DB 14; Length 1093;  
 Best Local Similarity 22.4%; Pred. No. 0.62;  
 Matches 37; Conservative 33; Mismatches 50; Indels 45; Gaps 5;  
  
 QY 3 NNDVAVLKRLKQGAADQIIEYLKQOVSILKE-----KAIQATLREKKLRVENAKL 57  
 DB 571 HNSNIITKRLRAKDKNENNVAKLNKKVLEELQHLKQVLOGKEVEKQRENIKKLNS 630  
 QY 58 EIEELKQEL-----IOAEIQNGVKQIAFPSTPLHANSVSENVIOSTA 101  
 DB 631 MYERQEKDLGRLOVMDLEEKNRISIQALDSEYKEL-----TDLAKAANAADSEKQDEAA 685  
 QY 102 VYTVSSGTKEQIKGTDGDKAKK-----IEKGEKKEKQOQISA 142  
 DB 686 LSR-----EMKAKKELSALEKQAEAFQOQETLA 715  
  
 RESULT 11  
 ABU70418  
 ID ABU70418 standard; Protein; 1122 AA.  
 XX  
 AC ABU70418;  
 XX  
 DT 10-JUN-2003 (first entry)  
 DT  
 XX  
 DE Human adipocyte Selected Interacting domain; SID, #49.  
 XX  
 KW Human; prey; adipocyte; SID; selected interacting domain;  
 KW anorectic; antidiabetic; protein-protein interaction; diabetes;  
 KW yeast 2-hybrid assay; metabolic disorder; obesity.  
 XX  
 OS Homo sapiens.  
 XX  
 PN MO200286122-A2.  
 XX  
 PD 31-OCT-2002.  
 XX  
 PF 14-MAR-2002; 2002MO-EP03768.

```

XX 14-MAR-2001; 2001US-275734P.
PR (HYBR-) HYBRIGENICS.
PA Legrain P, Daviet L;
XX WPI; 2003-103412/09.
DR N-PSDB; ACN56962.
XX

PT New complex between two interacting proteins in adipocyte cells, useful
PT for identifying selected interacting domains that modulate protein
PT interactions, or for preventing or treating metabolic disorders such as
PT obesity or diabetes
PS
PS Claim 6; Page 127-128; 382pp; English.
XX
XX The invention relates to a complex between two interacting proteins in
XX adipocyte cells, given in the specification. The proteins are identified
XX by selecting a bait protein from a known adipocyte marker and then
XX performing a yeast 2-hybrid selection to isolate prey proteins encoded by
XX members of an adipocyte cDNA library. The proteins are designated SID
XX (RTM) (selected interacting domains) proteins. Also included are a
XX polynucleotide encoding a polypeptide in the adipocyte cells, a
XX recombinant host cell expressing at least one of the interacting
XX polypeptides of the complex, selecting a modulating compound in adipocyte
XX cells, a SID (RTM) polypeptide comprising any of the 738 amino acid
XX sequences given in the specification (including its fragment or variant),
XX a SID (RTM) polynucleotide comprising any of the 738 nucleotide sequences
XX given in the specification (including its fragment or variant), a vector
XX comprising the SID (RTM) polynucleotide, a recombinant host cell
XX comprising the vector, a protein chip comprising the polypeptides and
XX a record comprising all or part of the data, listed in the specification.
XX The complex, polypeptides, polynucleotides and compounds are
XX useful for preventing or treating metabolic disorders such as obesity
XX or diabetes. The polynucleotides are useful as probes or primers. The
XX complex is particularly useful for identifying selected interacting
XX domains (SID (RTM)) for screening drugs that modulate the protein
XX interaction, thus exhibiting the therapeutic effect. The present
XX sequence represents a SID (prey) protein of the invention.
XX
SQ Sequence 1122 AA;
XX
XX Query Match 14.9%; Score 105.5; DB 24; Length 1122;
XX Best Local Similarity 26.1%; Pred. No. 0.79;
XX Matches 41; Conservative 30; Mismatches 55; Indels 31; Gaps 6;
OY 16 AADQIIIVLKQVSLKKEK-----AIIQATLREK-----KLRYENAKIKKEIEBLK 63
DB 520 SENKSIIEKLKDVISMNASSEFEVQIALNEAKLSEKVKSECHRVQENARLKKKKQLQ 579
OY 64 QEL-----IQAEIONGVK-----QIAPSGTPIHANSVSNVIGSTAVTTVSSGT 109
DB 580 QEIDWSKTHAEHLSQIKSFPEKSKQDLVAL---THKDNINALTNCITQALNLECESES 636
OY 110 KEQIKGTGDEKAKIEKKEKKEKQOQSIAGSAD 146
DB 637 EGQNGKGNDSPELANGEV--GGRNREKMKQIKQMD 671
XX
XX RESULT 12
XX AAM25602
XX ID AAM25602 standard; Protein; 1193 AA.
XX
XX AAM25602;
XX
XX 16-OCT-2001 (first entry)
XX
XX Human protein sequence SEQ ID NO:1117.
XX
XX Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
XX antiinflammatory; antipneumatic; antiautistic; immunosuppressive;
XX antibacterial; endocrine; cardiant; central nervous system; virucide;

```

```

KM anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;
KM antiaggregant; haemostatic; vulnary; antilucer; osteopathic; eczema;
KM dermatological; antiallergic; antiaesthetic; antidiabetic; cytostatic;
KM neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
KM immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
KM antiaphyllactic; rheumatoid arthritis; septic shock; pancreatitis;
KM cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
KM genetic disease; haematopoietic disorder; platelet disorder; asthma;
KM thrombocytopenia; osteoporosis; severe combined immunodeficiency;
KM allergic rhinitis; diabetes; multiple sclerosis; depression;
KM Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
KM neurological disorder.
XX
XX Homo sapiens.
XX
XX WO200153455-A2.
XX
XX 26-JUL-2001.
XX
XX 22-DEC-2000; 2000WO-US35017.
XX
XX 23-DEC-1999; 99US-0471275.
XX 21-JAN-2000; 2000US-0488725.
XX 25-APR-2000; 2000US-0552317.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Dmanac RT;
XX WPI; 2001-457603/49.
XX N-PSDB; AAH99543.
XX
XX Isolated human polynucleotides encoding polypeptides, useful for the
XX treatment and diagnosis of e.g. cancer, ulcers and HIV infection -
XX Claim 20; Page 230; 121pp; English.
XX
XX AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
XX AAM25963. The proteins can have activities based on the tissues and
XX cells they are expressed in, such as: antiinflammatory; antirheumatic;
XX antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;
XX central nervous system; virucide; anti-HIV; fungicide; antimutagen;
XX cardiovascular; antianaemic; antiaggregant; haemostatic; vulnary;
XX antifucier; osteopathic; dermatological; antiallergic; antiaesthetic;
XX antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
XX antiparkinsonian; and immunostimulant. The proteins and polynucleotides
XX encoding them can be used in gene therapy, antisense therapy and vaccine
XX production. The proteins and polynucleotides are useful for screening for
XX agonists or antagonists of a protein and for the treatment and diagnosis
XX of disorders associated with the activity of a protein e.g. inflammation,
XX rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
XX neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
XX infections, autoimmunity, genetic diseases, haematopoietic disorders,
XX anaemia, platelet disorders, thrombocytopenia, wounds, burns, ulcers,
XX osteoporosis, severe combined immunodeficiency, eczema, allergic
XX rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
XX Alzheimer's disease, Parkinson's disease, neurodegenerative and
XX neurological disorders.
XX
XX Sequence 1193 AA;
XX
XX Query Match 14.9%; Score 105.5; DB 22; Length 1193;
XX Best Local Similarity 26.1%; Pred. No. 0.86;
XX Matches 41; Conservative 30; Mismatches 55; Indels 31; Gaps 6;
OY 16 AADQIIIVLKQVSLKKEK-----AIIQATLREK-----KLRYENAKIKKEIEBLK 63
DB 591 SENKSIIEKLKDVISMNASSEFEVQIALNEAKLSEKVKSECHRVQENARLKKKKQLQ 650
OY 64 QEL-----IQAEIONGVK-----QIAPSGTPIHANSVSNVIGSTAVTTVSSGT 109
DB 651 QEIDWSKTHAEHLSQIKSFPEKSKQDLVAL---THKDNINALTNCITQALNLECESES 707

```



QY 16 AEAADQIIEYLKQVSLKEK-----ALQNTLREBK-----KLRVENAKLKEIEELK 63  
 DB 591 SENKSIIEKADVISMNASFSEVQIALNEAKLSEKSKSCHVQOEENARLKKKKQQLQ 650  
 QY 64 QEL-----IOELINGVK-----QIAPSGTPLHANSWSENVISQTAVTTSVSGT 109  
 DB 651 QEIDWSKTLHAELSEQISFEKSKQDLFVAL--THKDDNINALTCTQNLLECSSES 707  
 QY 110 KEQIKGGTGDEKKAKEKKEKKEKKEKQOSIAGSAD 146  
 DB 708 EGQNKGGNDSDDELANGEV--CGDRNEKKNQIKQMD 742

RESULT 15  
 ABB71141  
 ID ABB71141 standard; Protein, 1833 AA.

XX AC ABB71141;  
 XX DT 26-MAR-2002 (first entry)  
 XX DE Drosophila melanogaster polypeptide SEQ ID NO 40215.  
 XX KM Drosophila; developmental biology; cell signalling; insecticide;  
 XX KM pharmaceutical.  
 XX OS Drosophila melanogaster.  
 XX PN WO200171042-A2.  
 XX PD 27-SEP-2001.  
 XX PF 23-MAR-2001; 2001WO-US09231.  
 XX PR 23-MAR-2000; 2000US-191637P.  
 XX PR 11-JUL-2000; 2000US-0614150.  
 XX PA (PEKE ) PE CORP NY.  
 XX PI Venter JC, Adams M, Li FWD, Myers EW;  
 XX DR WPI; 2001-656860/75.  
 XX DR N-PSDB; ABL15244.  
 XX PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 XX PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 XX PT interactions -  
 XX PS Disclosure; SEQ ID NO 40215; 21pp + Sequence listing; English.  
 XX CC The invention relates to an isolated nucleic acid detection reagent  
 XX CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 XX CC useful in developmental biology and in elucidating cell signalling and  
 XX CC cell-cell interactions in higher eukaryotes for the development of  
 XX CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 XX CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA  
 XX CC sequences (AB101840-AB16175) and the encoded proteins  
 XX CC (AB57737-AB572072).  
 XX CC The sequence data for this patent did not form part of the printed  
 XX CC specification, but was obtained in electronic format directly from WIPO  
 XX CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX SQ Sequence 1833 AA;

Query Match 14.9%; Score 105.5; DB 22; Length 1833;  
 Best Local Similarity 27.0%; Pred. No. 1.5;  
 Matches 41; Conservative 24; Mismatches 56; Indels 31; Gaps 5;

QY 12 EQKGAEDQIIEYLKQVSLKEKAIQATIREBKLRVENAKLKEIEELKQELIQARI 71  
 DB 725 EDDPAEIRILELNEQESIRLKV-----EDLEKENAESKRYVRELQAKLRQ-DS 774  
 QY 72 QNGVKQIAPSGTPLHAN---SNVSENVISQTAVTTSVSGTKEQIKG-----GTGD 119

DB 775 SNGSKSILSLGTSSSAEKKVKTILNEELVQLRRTLTEKEQTVDSLKNQLSKLDLTLETEN 834  
 QY 120 EKAKEK-----TEKKEKKEKQOSIA 142  
 DB 835 DKLAKEKRLILAKKASEKTGEVDQKKESLA 866

Search completed: January 2, 2004, 18:20:30  
 Job time : 45.818 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2004, 18:17:12 ; Search time 11.5294 Seconds

(without alignments)  
599.590 Million cell updates/sec

Title: US-09-930-169-1

Perfect score: 709  
Sequence: 1 MANNDAVLKLEKQKGAEDQ.....EKKGKKKKQKQSIAGSADS 147

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database: SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	709	100.0	312	1	MCAL_HUMAN
2	598.5	77.4	359	1	MCAL_CRICK
3	509	71.8	310	1	MCAL_MOUSE
4	125.5	17.7	516	1	P54_EWTPC
5	117	16.5	662	1	TLPB_BACSV
6	116.5	16.4	1079	1	IF2P_SCHPO
7	108.5	15.3	662	1	MCBP_BACSV
8	106.5	15.0	1093	1	TWFI_HUMAN
9	106	15.0	461	1	US45_LACIC
10	104	14.7	473	1	YVCE_BACSV
11	104	14.7	570	1	STIM_DROME
12	104	14.7	961	1	UDP_BOVIN
13	103.5	14.6	561	1	EZPA_BACHD
14	103	14.5	1130	1	Y117_CABEL
15	103	14.5	1169	1	SMC_METUA
16	103	14.5	2230	1	GOO4_HUMAN
17	102	14.4	886	1	RA50_SITAC
18	102	14.4	1220	1	IF2P_HUMAN
19	101.5	14.3	539	1	MY33_HYDAT
20	101.5	14.3	966	1	H1PR_HUMAN
21	101	14.2	1068	1	KF5C_MOUSE
22	101	14.2	957	1	KF5C_HUMAN
23	101	14.2	1938	1	MY33_HUMAN
24	100.5	14.2	1690	1	CI90_DROME
25	100	14.1	1243	1	SMC4_MICAR
26	99.5	14.0	886	1	RA50_ARCVU
27	99.5	14.0	2116	1	MY32_DICDI
28	99	14.0	1940	1	MY33_CHICK
29	98.5	13.9	407	1	Y173_MOUSE
30	98.5	13.9	879	1	RA50_SUTLO
31	98	13.8	919	1	INCE_HUMAN
32	98	13.8	1197	1	YOIK_ECOLI
33	97.5	13.8	553	1	YOIK_ECOLI

34	97.5	13.8	727	1	MFPI_ARATH	Q0185 arabidopsis
35	97	13.7	1605	1	RRBI_MOUSE	Q9915 mus musculus
36	97	13.7	2434	1	YCP1_OENHO	Q0510 oenothera h
37	96.5	13.6	843	1	YMS1_YEAST	Q0505 saccharomyc
38	96	13.5	856	1	CLPB_HELPY	P71404 helicobacte
39	96	13.5	1526	1	MY32_SCHPO	Q0516 schizosacch
40	96	13.5	1534	1	RRBI_CANPA	Q28298 canis fam11
41	95.5	13.5	793	1	CALD_HUMAN	Q05682 homo sapien
42	95.5	13.5	1084	1	MY33_RABIT	P02562 oryctolagus
43	95.5	13.5	1447	1	SGS1_YEAST	P35187 saccharomyc
44	95.5	13.5	3214	1	BPA1_HUMAN	Q03001 homo sapien
45	95	13.4	579	1	G160_HUMAN	Q08378 homo sapien

## ALIGNMENTS

RESULT 1  
ID MCAL\_HUMAN STANDARD; PRT; 312 AA.  
AC Q12904; Q96CQ9;  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Multisynthetase complex auxiliary component p43 [Contains:  
DE Inducible cyclokinase activating polypeptide II (EMAP-II) (Small  
DE Inducible cytokine subfamily B member 1)].  
CN SCFEL OR EMAP2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
OX NCBI\_Taxid=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95014290; PubMed=7929199;  
RA Kao J.-I, Houck K., Fan Y., Haehnel I., Libutti S.K., Kayton M.L.,  
RA Glicksberg T., Chabot J., Nowyrod R., Greenberg S., Kuang W.J.,  
RA Leung D.W., Hayward J.R., Kiesel W., Heath M., Brett J., Stern D.M.,  
RT "Characterization of a novel tumor-derived cytokine. Endothelial-  
RT monocytic activating polypeptide II,"  
RL J. Biol. Chem. 269:25106-25119(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stepien L., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Umed T.B., Tomihata S., Carninci P., Frange C.,  
RA Bock S.A., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Roark S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hultk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fanev J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,  
RA Schermer A., Schein J.E., Jones S.J.M., Marra M.A.,  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences,"  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
CC 1- SIMILARITY: Contains 1 tRNA-binding domain.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and that statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>)

```

CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U0117; AAA62202.1; -.
CC DR EMBL: BC014051; AAH14051.1; -.
CC DR PDB: 1E7Z; 06-FEB-01.
CC DR PDB: 1EUJ; 06-SEP-00.
CC DR PDB: 1FLD; 07-FEB-01.
CC DR Genew: HGNC:10648; SCYE1.
CC DR MIM: 603605; -.
CC DR GO: GO:0005615; C:extracellular space; TAS.
CC DR GO: GO:0005125; F:cytokine activity; TAS.
CC DR GO: GO:0000049; F:RNA binding activity; TAS.
CC DR GO: GO:0006418; P:amino acid activation; TAS.
CC DR GO: GO:0006935; P:chemotaxis; TAS.
CC DR GO: GO:0006954; P:inflammatory response; TAS.
CC DR InterPro: IPR002547; tRNA_bind.
CC DR Pfam: PF01588; tRNA_bind; 1.
CC DR PIRSF: PIRSF005381; EMAP11; 1.
CC DR PROSITE: PS50886; TRBD; 1.
CC DR Protein biosynthesis; RNA-binding; tRNA-binding; Cytokine;
CC 3D-structure.
CC FT PROPR 1 146
CC FT CHAIN 147 312
CC FT DOMAIN 151 252
CC FT CONFLICT 79 79
CC FT SEQUENCE 312 AA; 34326 MW; 946310A0216F7587 CRC64;

Query Match 100.0%; Score 709; DB 1; Length 312;
Best Local Similarity 100.0%; Pred. No. 3.2e-38;
Matches 147; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MANNDVAVLRLEKRGKGAADQIIIEYKQVSLKKEKAILOATLREKKLRVENAKLKEIE 60
DB 1 MANNDVAVLRLEKRGKGAADQIIIEYKQVSLKKEKAILOATLREKKLRVENAKLKEIE 60

QY 61 ELKQELIQAEIQNGVQKQIAFP--SGTPPLHANSMSENVIOSTAVTVSSGTEQIKGTCG 120
DB 61 ELKQELIQAEIQNGVQKQIAFP--SGTPPLHANSMSENVIOSTAVTVSSGTEQIKGTCG 120

QY 121 KKAKEKIEKKGEKKKKQKQSIAGSADS 147
DB 121 KKAKEKIEKKGEKKKKQKQSIAGSADS 147

RESULT 2
MCAL_CRIGR STANDARD; PRT; 359 AA.
ID MCAL_CRIGR STANDARD; PRT; 359 AA.
AC 054873;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Multisynthetase complex auxiliary component p43 [Contains:
DE Endothelial-monocyte activating polypeptide II (EMAP-II) (Small
DE Inducible cytokine subfamily E member 1)].
GN SCYE1.
OS Cricetus griseus (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetulus.
OC NCBI_TaxID=10029;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98070438; PubMed=9405472;
RA Quesvillon S., Agou F., Robinson J.-C., Miranda M.;
RT "The p43 component of the mammalian multi-synthetase complex is likely
RT to be the precursor of the endothelial monocyte-activating polypeptide
RT II cytokine."
RL J. Biol. Chem. 272:32573-32579 (1997).
CC -1- SUBUNIT: COMPONENT OF THE MULTISYNTHETASE COMPLEX WHICH IS
CC COMPRISED OF A BIFUNCTIONAL GLUTAMYL-PROLYL-TRNA SYNTHETASE, THE
CC MONOSPECIFIC ISOLEUCYL, LEUCYL, GLUTAMINYL, METHIONYL, LYSYL,
CC ARGINYL, AND ASPARTYL-TRNA SYNTHETASES AS WELL AS THREE AUXILIARY

```

```

CC PROTEINS, P18, P48 AND P43.
CC -1- SIMILARITY: Contains 1 tRNA-binding domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF021800; AAB95207.1; -.
CC DR InterPro: IPR002547; tRNA_bind.
CC DR Pfam: PF01588; tRNA_bind; 1.
CC DR PIRSF: PIRSF005381; EMAP11; 1.
CC DR PROSITE: PS50886; TRBD; 1.
CC DR Protein biosynthesis; RNA-binding; tRNA-binding; Cytokine.
CC FT DOMAIN 198 299
CC FT SEQUENCE 359 AA; 39601 MW; 4D868D1B65D72C23 CRC64;

Query Match 77.4%; Score 548.5; DB 1; Length 359;
Best Local Similarity 79.9%; Pred. No. 4.6e-28;
Matches 119; Conservative 10; Mismatches 17; Indels 3; Gaps 2;

QY 1 MANNDVAVLRLEKRGKGAADQIIIEYKQVSLKKEKAILOATLREKKLRVENAKLKEIE 60
DB 47 MATNDVAVLRLEKRGKGAADQIIIEYKQVALLKKEKAVLOATLREKKLRVENAKLKEIE 106

QY 61 ELKQELIQAEIQNGVQKQIAFP--SGTPPLHANSMSENVIOSTAVTVSSGTEQIKGTCG 118
DB 107 ELKQELIQAEIQNGVQKQIPVPOSDTPVOASASVSTVISTISCSIKHSKG-G 165

QY 119 DEKKAKEKIEKKGEKKKKQKQSIAGSADS 147
DB 166 DEKKAKEKTDKKGKKEKKLQSAAPSADS 194

RESULT 3
MCAL_MOUSE STANDARD; PRT; 310 AA.
ID MCAL_MOUSE STANDARD; PRT; 310 AA.
AC P31230; Q60659;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Multisynthetase complex auxiliary component p43 [Contains:
DE Endothelial-monocyte activating polypeptide II (EMAP-II) (Small
DE Inducible cytokine subfamily E member 1)].
GN SCYE1 OR EMAP2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95014290; PubMed=7929199;
RA Kao J., Houck K., Fan Y., Haehnel I., Liputci S.K., Kayton M.L.,
RA Grikscheit T., Chabot J.R., Nowyrod R., Greenberg S., Kuang W.J.,
RA Leung D.W., Hayward J.R., Kistiel W., Heath M., Brett J., Stern D.M.;
RT "Characterization of a novel tumor-derived cytokine. Endothelial-
RT monocyte activating polypeptide II."
RL J. Biol. Chem. 269:25106-25119 (1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Scheffer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.D., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loughellano N.A., Peters G.J., Abramson R.D., Millhys S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

```

Db	119	EKKVKEKTEKKGEKKE-KQGSAA	STDS	145
RESULT 4				
ID	P54_ENTFC	STANDARD;	PRT;	516 AA.
AC	P13692;			
DT	01-JAN-1990 (Rel. 13, Created)			
DT	01-NOV-1991 (Rel. 20, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	P54 protein precursor.			
OS	Enterococcus faecium (Streptococcus faecium).			
OC	Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.			
OK	NCBI_taxonomy=1352;			
RM				
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=89385998; PubMed=2780297;			
RA	Fuerst P., Moesch H.-U., Solioz M.;			
RT	"A protein of unusual composition from Enterococcus faecium.";			
RL	Nucleic Acids Res. 17:6724-6724(1989).			
CC	-1- SUBCELLULAR LOCATION: Attached to the cell wall.			
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C40.			
CC				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb.ch/announce/">http://www.isb.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC				
DR	EMBL; X16421; CAA34442.1; ALT_INIT.			
DR	PIR; S05542; S05542.			
DR	MEROPS; C40; UPW; -.			
DR	InterPro; IPR000064; NRPC_P60.			
DR	Pfam; PF00877; NRPC_P60; I.			
KW	Signal; Cell wall.			
FT	SIGNAL 1 27	POTENTIAL.		
FT	CHAIN 28 516	P54 PROTEIN.		
FT	ACT SITE 429 429	POTENTIAL.		
SQ	SEQUENCE 516 AA; 54596 MM; 402ECMA439846D26 CRC64;			
Query Match	17.7%;	Score 125.5;	DB 1;	Length 516;
Best Local Similarity	28.1%;	Pred. No. 0.27;		
Matches	41;	Conservative 39;	Mismatches 51;	Indels 15;
			Gaps 5;	
QY	5 DAVLAKREFGKAEADQIIETIKQOVSLIKKAKIIQATLRBSKLRVENAKIKKEIEIKQ 64			
DB	38 DKKIADLQNOQASQSQSIEALBEGVASINTKA--QDLITRKQDITRKESAQIKQEIQLQ 95			
QY	65 --ELIQAEIONGVQKQAFPSGTP-----LHANSMSVENIYSTAVVTYVSSTGEIQK 116			
DB	96 RIERREKRTIQKARETQVQKNTSYDAVLNADSL--ADAVGRIOAMSTVIKANODLQV-- 152			
QY	117 TGEDEKAKKEIKKKEKKKKQKQKQK 142			
DB	153 --QKEDKQAVKAKKEAKKEKQKELA 176			
RESULT 5				
ID	TLPB_BACSU	STANDARD;	PRT;	662 AA.
AC	P39217;			
DT	01-FEB-1995 (Rel. 31, Created)			
DT	01-FEB-1995 (Rel. 31, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Methyl-accepting chemotaxis protein tlpB.			
GN	TLPB.			
OS	Bacillus subtilis.			
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.			
OK	NCBI_taxonomy=1423;			
RM				

RP SEQUENCE FROM N.A.  
 RC STRAIN=168 / OI1085;  
 RX MEDLINE=94245722; PubMed=8188684;  
 RA Hanlon D.W., Ordal G.W.;  
 RT "Cloning and characterization of genes encoding methyl-accepting  
 RT chemotaxis proteins in *Bacillus subtilis*,"  
 RL J. Biol. Chem. 269:14038-14046(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=98044033; PubMed=9384377;  
 RA Kunst F., Ogassawara N., Moszer I., Albertini A.M., Alloni G.,  
 RA Azevedo V., Bertoletti M.G., Beesleres P., Bolocin A., Borchert S.,  
 RA Boursier R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
 RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,  
 RA Denicof F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
 RA Entian K.D., Errington J., Fabret C., Ferrati E., Foulger D.,  
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 RA Ghim S.Y., Glaeser P., Goffeau A., Golightly E.J., Grandi G.,  
 RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,  
 RA Joris B., Karamata D., Kasaahara Y., Klaert-Blanchard M., Klein C.,  
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
 RA Parro V., Pohl T.M., Portetelle D., Potwolik S., Prescott A.M.,  
 RA Pressac E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 RA Rieger M., Rivolta C., Roche B., Rose M., Sadie Y.,  
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
 RA Sekiguchi J., Sekowska A., Seror S.J., Serron P., Shin B.S., Soldo B.,  
 RA Sorkin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
 RA Takeuchi M., Takemoto A., Tanaka T., Terpestra P., Tognoni A.,  
 RA Tosafo V., Uchiyama S., Vandenbol M., Vannier F., Vassartoli A.,  
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenger T.,  
 RA Winters K., Wipert A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;  
 RT "The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*,"  
 RT Nature 390:249-256(1997).  
 RL Nature 390:249-256(1997).  
 CC -1- FUNCTION: CHEMOTACTIC-SIGNAL TRANSDUCERS RESPOND TO CHANGES IN THE  
 CC CONCENTRATION OF ATTRACTANTS AND REPELLENTS IN THE ENVIRONMENT,  
 CC TRANSDUCE A SIGNAL FROM THE OUTSIDE OF THE CELL, AND  
 CC FACILITATE SENSORY ADAPTATION THROUGH THE VARIATION OF THE LEVEL  
 CC OF METHYLATION. ALL AMINO ACIDS SERVE AS ATTRACTANTS IN  
 CC B. SUBTILIS, THEY APPEAR TO CAUSE AN INCREASE IN THE TURNOVER  
 CC METHYL GROUPS, LEADING TO METHYLATION OF AN UNIDENTIFIED ACCEPTOR,  
 CC WHILE REPELLENTS HAVE BEEN SHOWN TO CAUSE A DECREASE IN METHYL.  
 CC GROUP TURNOVER. THE METHYL GROUPS ARE ADDED BY A METHYLTRANSFERASE  
 CC AND REMOVED BY A METHYLESTERASE.  
 CC -1- SIMILARITY: Contains 1 methyl-accepting transducer domain.  
 CC -1- SIMILARITY: Contains 1 HAMF domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL custodian -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: L29189; AAA20557.1; -  
 CC EMBL: 299119; CAB15101.1; -  
 CC EMBL: 299120; CAB15112.1; -  
 CC PIR: D54078; D54078.  
 CC HSSP: P02942; 10U7.  
 CC Subtilin; Bg10862; tlpB.  
 CC InterPro: IPR004010; Cache.  
 CC InterPro: IPR004089; Chemtaxis\_transd.  
 CC InterPro: IPR003660; HAMF.  
 CC InterPro: IPR003123; TarH.

DR Pfam; PF02743; Cache, 1.  
 DR Pfam; PF00672; HAMF, 1.  
 DR Pfam; PF00015; MCPSignal, 1.  
 DR Pfam; PF02203; TarH, 1.  
 DR SMART; SM00304; HAMF, 1.  
 DR SMART; SM00283; NA, 1.  
 DR SMART; SM00319; TarH, 1.  
 DR PROSITE; PS00111; CHEMOTAXIS\_TRANSDUC\_2, 1.  
 DR PROSITE; PS00885; HAMF, 1.  
 KM Chemotaxis; Transducer; Transmembrane; Methylation; Complete proteome.  
 FT DOMAIN 1 16  
 FT TRANSMEM 17 37  
 FT DOMAIN 38 281  
 FT TRANSMEM 282 302  
 FT DOMAIN 303 662  
 FT DOMAIN 303 355  
 FT DOMAIN 374 610  
 FT MOD\_RES 370 370  
 FT MOD\_RES 594 594  
 FT MOD\_RES 629 629  
 FT MOD\_RES 636 636  
 FT MOD\_RES 662 662  
 SQ SEQUENCE 662 AA; 71535 NM; 91215f86293D7425 CRC64;  
 Query Match 16.5%; Score 117; DB 1; Length 662;  
 Best local Similarity 26.0%; Pred. No. 1.2;  
 Matches 45; Conservative 26; Mismatches 64; Indels 38; Gaps 5;  
 QY 5 DAVLKRLKQKADADITLYK--QVSLKELKILQ-----TLREKLR 49  
 DB 463 EAVVKLEKFSDDITSLVINGIDQNTLLALNAIEARAGEYGRGFSVAEEVRKLA 522  
 QY 50 VNAKTKKEIEELKKEILQ-----AEIONGYKQIAFPSCGTPPLHANSVENVYST 100  
 DB 523 VQSAASAKIEBGLDIEIVAEITSLSMFQSVHAYKE-----GLQITDQIAESKQLY 575  
 QY 101 AVTVSSG-----TKEDIKGTDEKKAKEIKKKEKKKQKQSGASD 146  
 DB 576 EMTTQISGLQLQNLNTVBEQLSAGQSVSSAVDISVAKESAGIQDIAMAE 628  
 RESULT 6  
 ID PF2P SCHPO STANDARD; PRT; 1079 AA.  
 AC 010251;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Probable translocation initiation factor Pf-2.  
 GN SPAC56F6.03.  
 OS Schizosaccharomyces pombe (Pisces yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 NX NCBI TaxId=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RX MEDLINE=21848401; PubMed=11859360;  
 RA Wood V., Gilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sgouras J., Peat N., Hayes J., Baker S., Basham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Frazer A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Wolcaert G., Aert R., Robben J., Grynponez B.,  
 RA Welfens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,

RX MEDLINE=94245722; PubMed=8188684;  
 RA Hanlon D.W., Ordal G.W.;  
 RT "Cloning and characterization of genes encoding methyl-accepting  
 RT chemotaxis proteins in *Bacillus subtilis*."  
 RL J. Biol. Chem. 269:14038-14046(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=98044033; PubMed=9384377;  
 RA Kunst F., Osasawara N., Moser I., Albertini A.M., Alloni G.,  
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
 RA Boriss R., Boursier L., Brans A., Brann M., Brigell S.C., Bron S.,  
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
 RA Choi S.K., Codani U.J., Conerton I.F., Cummings N.J., Daniel R.A.,  
 RA Denizot F., Devine K.J., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
 RA Ertian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 RA Ghm S.Y., Glaeser P., Goffeau A., Goughly E.J., Grandi G.,  
 RA Giuseppe G., Guy B.J., Haga K., Haeich U., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 RA Joris B., Karamata D., Kasehara Y., Kleier-Blanchard M., Klein C.,  
 RA Kobayashi Y., Koetter P., Koningslehn G., Krogh S., Kumano M.,  
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
 RA Lee S.W., Levine A., Liu H., Maeda S., Manuel C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno M., Mostl D., Nakai S., Noback M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogawara A., Oudega B., Park S.H.,  
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
 RA Pesecean E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 RA Rieger M., Rivolta C., Roche A., Roche B., Rose M., Sadle Y.,  
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
 RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,  
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
 RA Takenuchi M., Tanakoshi A., Tanaka T., Terpiltra P., Tognoni A.,  
 RA Terasaki Y., Uchiyama S., Vandenbol M., Vannier F., Vassart A.,  
 RA Viari A., Wamburt R., Wedler E., Wedler H., Weitzneger T.,  
 RA Winters P., Wipikt A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;  
 RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*  
 RT *subtilis*."  
 RL Nature 390:249-256(1997).  
 RL -I- FUNCTION: RESPONSIBLE FOR DETECTING A SUBSET OF AMINO ACIDS THAT  
 CC INCLUDES ASPARAGINE, ASPARTATE, GLUTAMINE, AND HISTIDINE.  
 CC -I- FUNCTION: CHEMOTACTIC-SIGNAL TRANSDUCERS RESPOND TO CHANGES IN THE  
 CC CONCENTRATION OF ATTRACTANTS AND REPELLENTS IN THE ENVIRONMENT,  
 CC TRANSDUCE A SIGNAL FROM THE OUTSIDE TO THE INSIDE OF THE CELL, AND  
 CC FACILITATE SENSORY ADAPTATION THROUGH THE VARIATION OF THE LEVEL  
 CC OF METHYLATION. ALL AMINO ACIDS SERVE AS ATTRACTANTS IN  
 CC B. SUBTILIS, THEY APPEAR TO CAUSE AN INCREASE IN THE TURNOVER  
 CC METHYL GROUPS, LEADING TO METHYLATION OF AN UNIDENTIFIED ACCEPTOR,  
 CC WHILE REPELLENTS HAVE BEEN SHOWN TO CAUSE A DECREASE IN METHYL  
 CC GROUP TURNOVER. THE METHYL GROUPS ARE ADDED BY A METHYLTRANSFERASE  
 CC AND REMOVED BY A METHYLESTERASE.  
 CC -I- SIMILARITY: Contains 1 methyl-accepting transducer domain.  
 CC -I- SIMILARITY: Contains 1 HAMP domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.1ab-61b.ch/announce/emb-1>).  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
 CC -----  
 DR EMBL; L29189; AAA20554.1; -;  
 DR EMBL; Z99119; CAB15104.1; -;  
 DR EMBL; Z99120; CAB15115.1; -;  
 DR PIR; A54078; A54078.  
 DR HSSP; P02942; 1QU7.  
 DR Subtilist; BG10859; mcpB.  
 DR Interpro; IPR004010; Cache.  
 DR Interpro; IPR004089; Chmtaxis\_transd.  
 DR Interpro; IPR003660; HAMP.  
 DR Interpro; IPR003122; TArH.

```

DR Pfam; PF02743; Cache; 1.
DR Pfam; PF00672; HAMP; 1.
DR Pfam; PF00015; MCPsignal; 1.
DR Pfam; PF02203; Tact; 1.
DR SMART; SM00304; HAMP; 1.
DR SMART; SM00283; MA; 1.
DR SMART; PF05011; CHEMOTAXIS_TRANSLOC_2; 1.
DR PROSITE; PS00119; CHEMOTAXIS_TRANSLOC_2; 1.
DR PROSITE; PS00885; HAMP; 1.
DR Chemotaxis; Transducer; Transmembrane; Methylation; Complete proteome.
KM DOMAIN 1 16 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 17 37 POTENTIAL.
FT DOMAIN 38 282 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 283 303 POTENTIAL.
FT DOMAIN 304 662 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 304 356 HAMP.
FT DOMAIN 375 611 METHYL-ACCEPTING TRANSDUCER.
FT MOD_RES 371 371 DEAMIDATION AND METHYLATION (BY SIMILARITY).
FT MOD_RES 595 595 DEAMIDATION AND METHYLATION (BY SIMILARITY).
FT MOD_RES 630 630 METHYLATION (BY SIMILARITY).
FT MOD_RES 637 637 METHYLATION (BY SIMILARITY).
SQ SEQUENCE 662 AA; 71885 MW; 98DD181FB46900BC CRC64;

Query Match 15.3%; Score 108.5; DB 1; Length 662;
Best Local Similarity 26.4%; Pred. No. 4.1;
Matches 44; Conservative 27; Mismatches 56; Indels 37; Gaps 7;

QY 5 DAVLKRLKQKGAADQIIEYLK---QVSLKKEKRLKLA-----TLREKKLR 49
DB 464 EAVVKGLEKSKDITSLRVINGIADQTNLALNAALRAGESGRGFSVAEEVKLA 523
QY 50 VENAKLKEETLEKQELQIOLQVQKQIAFPSTGPLHANSMSENVIOSTATTYSSGT 109
DB 524 VQSDASAKELIKELQIV-ARID-----TSIHMKRYVQEV-QSGLV--VTDDT 568
QY 110 KEQIKG-----GTGDEKKAKKEIKKKKKKKQKQSSAD 146
DB 569 KESFQSFMTNIEAGLQTMNSTVEQLSPRSQHVSAVSGIAD 612

RESULT 8
TMFL_HUMAN STANDARD; PRT; 1093 AA.
AC P82054;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE TATA element modulatory factor (TMF).
GN TMFL.
OS Homo sapiens (human).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A., AND FUNCTION.
RC TISSUE=Cervical carcinoma;
RX MEDLINE=93028466; Pubmed=1409643;
RA Garcia J.A., Ou S.-H.T., Wu F., Lusis A.J., Sparkes R.S., Gaynor R.B.;
RT "Cloning and chromosomal mapping of a human immunodeficiency virus 1
RT 'TAR' element modulatory factor.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:9372-9376(1992).
CC -!- FUNCTION: THIS PROTEIN BINDS THE HIV-1 TATA ELEMENT AND INHIBITS
CC TRANSCRIPTIONAL ACTIVATION BY THE TATA-BINDING PROTEIN (TBP).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC EMBL; L01042; AAD54608.1; -.
DR PIR; A47212; A47212.
DR TRANSFAC; T00835; -.
DR Genew; HGNC; 11870; TMFL.
DR MIM; 601126; -.
DR GO; GO:0003702; F:RNA polymerase II transcription factor acti. .; TAS.
DR GO; GO:0003712; F:transcription cofactor activity; TAS.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; TAS.
DR GO; GO:0006366; P:transcription from Pol II promoter; TAS.
KM Transcription regulation; DNA-binding; Repressor; Coiled coil.
FT DOMAIN 439 922 COILED COIL (POTENTIAL).
FT DOMAIN 984 1092 COILED COIL (POTENTIAL).
SQ SEQUENCE 1093 AA; 123170 MW; 26133E85F4677BE CRC64;

Query Match 15.0%; Score 106.5; DB 1; Length 1093;
Best Local Similarity 22.4%; Pred. No. 9.1;
Matches 37; Conservative 33; Mismatches 50; Indels 45; Gaps 5;

QY 3 NNDVAVKRLKQKGAADQIIEYLKQVSLKKE-----KALQATLREKKLRVENAKLK 57
DB 571 HNSNIIKKLRADQKENVVAKLKKVKELELQHLKQVLDQKEVEKQHRENIKKLNS 630
QY 58 ETEELKQEL-----IQAEIQVQKQIAFPSTGPLHANSMSENVIOSTATTYSSGT 101
DB 631 MYERQEKDILGRQVDMDELEKKRSIQALDSAYKEL-----TDLHKANAKDSEAOEAA 685
QY 102 VTTVSSGTKEQIKGNGDEKKAKKE-----IEKKKKKKKKQSSIA 142
DB 686 LSR-----EMKAKEELSAALDEKQAEARQOQETLA 715

RESULT 9
US45_LACIC STANDARD; PRT; 461 AA.
AC P22865;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Secreted 45 kDa protein precursor.
GN USP45.
OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1359;
[1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 28-37.
RC STRAIN=M61363;
RX MEDLINE=91071599; Pubmed=2123812;
RA van Asseldomk M., Rutten G., Oteman W., Siezen R.J., de Vos W.M.,
RA Simons G.;
RT "Cloning of usp45, a gene encoding a secreted protein from
RT Lactococcus lactis subsp. lactis M61363.";
RL Gene 95:155-160(1990).
CC -!- SIMILARITY: TO E.FABCTUM P54.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M60178; AAA5230.1; -.
DR EMBL; AL7083; CA01320.1; -.
DR Pfam; PF05257; AXE; 1.
KM Signal.
FT SIGNAL 1 27
FT CHAIN 28 461 SECRETED 45 kDa PROTEIN.
FT DOMAIN 299 314 POLY-SER.
SQ SEQUENCE 461 AA; 47387 MW; 51493C42224F3C03 CRC64;

Query Match 15.0%; Score 106; DB 1; Length 461;

```





RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballw R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolhakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies S.M.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.B., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evansgela C.C., Ferraz C., Ferriera S., Fleischmann W.,  
 RA Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glaeser K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howard T.J., Wei M.-H., Idegawa C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kenison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laoko P., Lai Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshirei A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RA "The genome sequence of *Drosophila melanogaster*.";  
 RA Science 287:2185-2195(2000).  
 RL [3]  
 RN REVISIONS.  
 RP STRAIN=Berkeley;  
 RC MEDLINE=22426069; PubMed=12537572;  
 RX Miera S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
 RA Hradecky P., Huang Y., Minkner J.S., Millburn G.H., Prochuk S.E.,  
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
 RA Battencourt B.R., Celisner S.B., de Grey A.D.N.J., Drysdale R.A.,  
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
 RA Lewis S.E.;  
 RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a  
 RT systematic review.";  
 RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkeley; TISSUE=Embryo;  
 RX MEDLINE=22426066; PubMed=12537569;  
 RA George R.A., Carlson J.W., Brokstein P., Yu C., Champe M.,  
 RA Stapleton M., Garin H., Krommiller B., Pacle J.M., Park S., Wan K.H.,  
 RA Rubin G.M., Celisner S.E.;  
 RT "A *Drosophila* full-length cDNA resource.";  
 RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).  
 CC -1- FUNCTION: Possible adhesion molecule (by similarity).  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Cell surface  
 CC (potential).  
 CC -1- SIMILARITY: Contains 1 sterile alpha motif (SAM) domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/ebis/announce/>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC -----  
 CC EMBL: AF228906; AAK82338.1; -  
 CC EMBL: AB003500; AAF48542.2; -  
 CC EMBL: AY069866; AAL39831.1; -  
 CC FlyBase; FBgn0045073; Stm.

DR InterPro; IPR001660; SAM.  
 DR Pfam; PF00536; SAM, 1.  
 DR SMART; SM00454; SAM, 1.  
 DR PROSITE; PS0105; SAM\_DOMAIN, 1.  
 KW Cell adhesion; Transmembrane; Coiled coil; Signal; Glycoprotein.  
 FT SIGNAL 1 23  
 FT CHAIN 24 570  
 FT DOMAIN 24 294  
 FT TRANSMEM 295 312  
 FT DOMAIN 313 570  
 FT DOMAIN 213 281  
 FT DOMAIN 26 95  
 FT DOMAIN 310 407  
 FT DOMAIN 420 462  
 FT CARBOHYD 212 212  
 SQ SEQUENCE 570 AA; 64797 MW; B8DC7917F379D0B5 CRC64;  
 Query Match 14.7%; Score 104; DB 1; Length 570;  
 Best Local Similarity 28.7%; Pred. No. 6.8;  
 Matches 43; Conservative 23; Mismatches 64; Indels 20; Gaps 4;  
 QY 1 MANNDAVLRRLROKGAADQIIIEYLK-QQVSLKKEKALIQATLRBEKKLRVENA-----K 54  
 DB 324 MWDNEGELPRAQSLQEMQKELERAMEQENVATELRLKKEAPTSSNSDLEVOQ 383  
 QY 55 LKKEIELKQELIOAEIQNGVQKQIAFPSCGPLHANSVSENVIOSTAVTVSSGTKEQIK 114  
 DB 384 LKKEIEMLNELNSRAFE-----LVNDCWSPPPQGLQVLYQTYELSSGNHQK 430  
 QY 115 GGTDEKAKAKETIKKGEKKKKQOSIAGS 144  
 DB 431 KRTSAEKQL-OSAREACEKLRKRSSTLVGA 459

RESULT 12  
 ID VDP\_BOVIN STANDARD; PRT; 961 AA.  
 AC P41541;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE General vesicular transport factor p115 (Transcytosis associated  
 DE protein) (TAP) (Vesicle docking protein).  
 GN VDP.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC TISSUE=Kidney.  
 RX MEDLINE=95132632; PubMed=7831323;  
 RA Sappenstein S.K., Walter D.M., Grosvener A.R., Heuser J.E.,  
 RA Waters M.G.;  
 RT "p115 is a general vesicular transport factor related to the yeast  
 RT endoplasmic reticulum to Golgi transport factor Usolp.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 92:522-526(1995).  
 RN [2]  
 RP CHARACTERIZATION.  
 RX MEDLINE=92381091; PubMed=1512287;  
 RA Waters M.G., Clary D.O., Rothman J.E.;  
 RT "A novel 115-kD peripheral membrane protein is required for  
 RT intercompartmental transport in the Golgi stack.";  
 RL J. Cell Biol. 118:1015-1026(1992).  
 CC -1- FUNCTION: GENERAL VESICULAR TRANSPORT FACTOR REQUIRED FOR  
 CC INTERCOMPARTMENTAL TRANSPORT IN THE GOLGI STACK; IT IS REQUIRED FOR  
 CC TRANSCYTOTIC FUSION AND/OR SUBSEQUENT BINDING OF THE VESICLES TO  
 CC THE TARGET MEMBRANE. MAY WELL ACT AS A VESICULAR ANCHOR BY  
 CC INTERACTING WITH THE TARGET MEMBRANE AND HOLDING THE VESICULAR  
 CC AND TARGET MEMBRANES IN PROXIMITY.  
 CC -1- SUBUNIT: HOMODIMER WITH TWO N-TERMINAL HEADS AND A C-TERMINAL  
 CC COILED-COIL TAIL. DIMER FORMED BY PARALLEL ASSOCIATION OF THE



CC TAILS.  
CC -1- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN WHICH RECYCLES  
CC BETWEEN THE CYTOSOL AND THE GOLGI APPARATUS DURING INTERPHASE.  
CC -1- DOMAIN: COMPOSED OF A GLOBULAR HEAD, AN ELONGATED TAIL (COILED-  
CC COIL) AND A HIGHLY ACIDIC C-TERMINAL DOMAIN.  
CC -1- PTM: PHOSPHORYLATED IN A CELL CYCLE-SPECIFIC MANNER;  
CC PHOSPHORYLATED IN INTERPHASE BUT NOT IN MITOTIC CELLS.  
CC DEPHOSPHORYLATED PROTEIN ASSOCIATES WITH THE GOLGI MEMBRANE;  
CC PHOSPHORYLATED PROMOTES DISSOCIATION (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE VDP/USO1/YBL047C FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL, U14186; AAA62631.1; -  
CC InterPro: IPR000225; Armadillo.  
CC InterPro: IPR006955; USO1\_p115\_C.  
CC InterPro: IPR006953; USO1\_p115\_head.  
CC Pfam: PF04871; USO1\_p115\_C\_1.  
CC Pfam: PF04869; USO1\_p115\_head\_1.  
CC PROSITE: PS50176; ARM\_REPEAT; UNKNOWN\_1.  
CC Transprot; Protein transport; Golgi stack; Membrane; Coiled coil;  
CC Phosphorylation.  
CC DOMAIN 1 641 GLOBULAR HEAD.  
CC FT DOMAIN 642 929 COILED COIL (POTENTIAL).  
CC FT DOMAIN 934 961 ASP/GLU-RICH (ACIDIC).  
CC FT MOD\_RES 941 941 PHOSPHORYLATION (BY SIMILARITY).  
CC SQ SEQUENCE 961 AA; 107514 MW; 993FEB7F90ABC0AC CRC64;  
Query Match 14.7%; Score 104; DB 1; Length 961;  
Best Local Similarity 23.3%; Pred. No. 11;  
Matches 42; Conservative 35; Mismatches 45; Indels 58; Gaps 7;  
QY 1 MANDAVLRLEQKGAADQIIEYLKQGVSLK-----EKALIQ--ATLBEKK---L 48  
DB 645 LEQHDSTVTHKMKIRBQDLQLEIKQIISTLKCONELQTAIVQGVSOIQOHQDQVYL 704  
QY 49 RV-----ENAKLKEIEBLK--QELIQEIQNGVQKQIAFPS 82  
DB 705 KVQGXGDSQHQGYTDGQNMNGVQPEISRLBEIEBLKSRBELLSQLAE----- 755  
QY 83 GTPHANSMSENVIOGTAVTVSSGKKEQIKGGTGD-EKQAKKEIKKKEKKKQOSI 141  
DB 756 -----KDSLEIKKSSQLSPGTWEOSSATAGSEQIAELKQELATLKQSQINSQSV 805  
RESULT 13  
EZRA\_BACHD STANDARD; PRT; 561 AA.  
ID EZRA\_BACHD Q9K802;  
AC Q9K802;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Separation ring formation regulator.  
GN EZRA OR BH3205.  
OS Bacillus halodurans.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=86665;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C-125 / JCM 9153;  
RX MEDLINE=20512582; PubMed=11058132;  
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,  
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kihara S.,  
RA Horikoshi K.;  
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus  
RT halodurans and genomic sequence comparison with Bacillus subtilis.",  
RL Nucleic Acids Res. 28:4317-4331(2000).

CC -1- FUNCTION: NEGATIVE REGULATOR OF FT23 RING FORMATION; MODULATES THE  
CC FREQUENCY AND POSITION OF FT23 RING FORMATION. INHIBITS FT23 RING  
CC FORMATION AT POLAR SITES. INTERACTS EITHER WITH FT23 OR WITH ONE  
CC OF ITS BINDING PARTNERS TO PROMOTE DEPOLYMERIZATION (BY  
CC SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: ASSOCIATED WITH THE CELL MEMBRANE AND ALSO  
CC COLOCALIZED WITH FT23 TO THE NASCENT SEPTAL SITE (BY SIMILARITY).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL, AP001518; BAB06924.1; -  
CC PIR; E84050; E84050.  
CC Cell division; Septation; Transmembrane; Coiled coil;  
CC Complete proteome.  
CC TRANSSEM 1 21 POTENTIAL.  
CC FT DOMAIN 94 133 COILED COIL (POTENTIAL).  
CC FT DOMAIN 194 214 COILED COIL (POTENTIAL).  
CC FT DOMAIN 251 466 COILED COIL (POTENTIAL).  
CC SQ SEQUENCE 561 AA; 65161 MW; C5C315CA960F0666 CRC64;  
Query Match 14.6%; Score 103.5; DB 1; Length 561;  
Best Local Similarity 26.4%; Pred. No. 7.2;  
Matches 47; Conservative 23; Mismatches 53; Indels 55; Gaps 6;  
QY 5 DAVLRLEQKGA-----ADQIIEYLKQGVSLKKEKALIQAT 41  
DB 254 DQWERLEEKRVLLBQLYLECGMEEINFTESNEQMFLEKKEVEKNETTILLPN 313  
QY 42 LRE-----EKKLR-----VENAKLKEIEBLKQEL-IQAEIQNGV 75  
DB 314 LREDLTTEELTLTKETESVQSYRLAEELVFOQKLEKLELQQLQVDEVTBEQ 373  
QY 76 KQIAFPGTPLHNSMSENVIOGTAVTVSSGKKEQIKGGTGD-EKQAKKEIKKKEKKK 133  
DB 374 KQ-TFSS-----VRSMLEWRREGITACNKRIEQAESLNSLRDELKAKELKQKER 425  
RESULT 14  
YL17\_CAEEL STANDARD; PRT; 1130 AA.  
ID YL17\_CAEEL Q11102;  
AC Q11102;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical 131.5 kDa protein C02F12.7 in chromosome X.  
GN C02F12.7.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Pelodierinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Miller N.;  
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
CC  
CC -1- SIMILARITY: WEAK, TO MYOSINS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL, U41545; AAK39135.1; -  
CC PIR; T34081; T34081.

DR WormRep; CO2F12.7; CE03901.  
 KW Hypothetical protein; Coiled coil.  
 FT DOMAIN 121 779 COILED COIL (POTENTIAL).  
 FT DOMAIN 805 1061 COILED COIL (POTENTIAL).  
 SQ SEQUENCE 1130 AA; 131485 MW; B0P2DFE3D99FB09 CRC64;  
 Query Match 14.5%; Score 103; DB 1; Length 1130;  
 Best Local Similarity 20.6%; Pred. No. 16;  
 Matches 37; Conservative 44; Mismatches 59; Indels 40; Gaps 5;  
 QY 4 NDAVLRLEQKGAADQIIIEYLK---QVSLKEKAILQATLREKKLRVENAKLKEE- 59  
 DB 870 DNSIOEKIKKEKATINEMERLKSRENELAKLHEEMYMOT--QNEKREBOSKLPQELM 927  
 QY 60 -----BELKQELIOAEIO-----NGVKQIAFPESGTPPLHANSWSEVNIQSTAV 102  
 DB 928 FEKEQLEAEKAEOSHLEAEVEQVFOADKESKKEQIEDLENLQKRELIOQLQDPTDB 987  
 QY 103 TTVSSGTEKQI-----KGGTDEKKAKEKIKKEKKKQOSIAGSA 145  
 DB 988 STEPHTKKMSITSHGVFNPFVSQMKDKKEASEKRTREAEKKAKEKEKAKAKA 1047  
 RESULT 15  
 SMC METJA STANDARD; PRT; 1169 AA.  
 AC 059037;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Chromosome partition protein smc homolog.  
 GN M1643.  
 OS Methanococcus jannaschii.  
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
 OC Methanocaldococcales; Methanocaldococcus.  
 OX NCBI\_TaxID=2190;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;  
 RX MEDLINE=96337999; PubMed=8688087;  
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,  
 RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,  
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,  
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrman J.L., Nguyen D.,  
 RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
 RA "Complete genome sequence of the methanogenic archaeon, Methanococcus  
 jannaschii";  
 RL Science 273:1058-1073(1996).  
 RN [2]  
 RP REVISIONS.  
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,  
 RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,  
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,  
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrman J.L., Nguyen D.,  
 RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
 RA Submitted (PEB-1998) to the EMBL/GenBank/DBJ databases.  
 RL - FUNCTION: PLAYS AN IMPORTANT ROLE IN CHROMOSOME STRUCTURE AND  
 PARTITIONING. ESSENTIAL FOR CHROMOSOME PARTITION (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE SMC FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL; U67604; AAB9663.1; -.  
 DR TIGR; M1643; -.  
 DR InterPro; IPR003439; ABC transporter.  
 DR InterPro; IPR005289; GTP-binding\_dom.  
 DR InterPro; IPR003405; SMC\_C.  
 DR InterPro; IPR003395; SMC\_N.  
 DR Pfam; PF02483; SMC\_C\_1.  
 DR Pfam; PF02463; SMC\_N\_1.  
 DR TIGRPFAM; TIGR00650; MG442; 1.  
 KW Hypothetical protein; ATP-binding; Coiled coil; Complete proteome.  
 FT NP\_BIND 31 38 ATP (POTENTIAL).  
 FT DOMAIN 160 521 COILED COIL (POTENTIAL).  
 FT DOMAIN 673 1032 COILED COIL (POTENTIAL).  
 SQ SEQUENCE 1169 AA; 136634 MW; B63CCE34E4C03F36 CRC64;  
 Query Match 14.5%; Score 103; DB 1; Length 1169;  
 Best Local Similarity 24.0%; Pred. No. 16;  
 Matches 35; Conservative 32; Mismatches 31; Indels 48; Gaps 6;  
 QY 3 NNDVLRLEQKGAADQIIIEYLK---QVSLKEKAILQATLREKKLRVENAKLKEE 58  
 DB 281 NN--IINELNKGNE--EVLHLKSIKELEVEIENDKVLDSINELKKVEIENKKKE 336  
 QY 59 IELKQELIOAEIONGVQIAFPESGTPPLHANSWSEVNIQSTAVTVSSGTEKQIKGTG 118  
 DB 337 IKETQKTIENR-----DSIIE-----KEQ----- 356  
 QY 119 DEKKAKEKKEKKEKKKQOSIAGS 144  
 DB 357 QIKELIEKIKNLNVEKELKEAIAES 382

Search completed: January 2, 2004, 18:21:03  
 Job time : 13.5294 secs



RESULT 2  
ID Q9KJ3 PRELIMINARY; PRT; 576 AA.  
AC Q9KJ3;  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
DE Secreted antigen SagB.  
GN SAGB.  
OS Enterococcus hirae.  
OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.  
RN NCBI\_TaxID=1354;  
[1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=IX2751;  
RA Teng F., Murray B.E., Weinstock G.M.;  
RT "Characterization of an enterococcal secreted antigen."  
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF167576; AAF87093.1; -  
DR InterPro: IPR000064; NLP\_C\_P60.  
DR Pfam: PF00877; NLP\_C\_P60; I.  
SQ SEQUENCE 576 AA; 60807 MW; 76F453E2780B57A7 CRC64;  
Query Match 17.7%; Score 125.5; DB 2; Length 576;  
Best Local Similarity 28.1%; Pred. No. 0.81; Mismatches 51; Indels 15; Gaps 5;  
Matches 41; Conservative 39; Mismatches 51; Indels 15; Gaps 5;  
QY 5 DAVLKRLQKGAENDQIIEYLKQVSLKKEKALIQATLREKKLRVENAKLKEIEELKQ 64  
DB 38 DKXITADLQNGQASQSGIEALLEGQVSAINTKA--QDLITQDRLTKRSADLKEIDLQ 95  
QY 65 --ELIQAEIQGVKQIAFPESGTP-----LHANSWSENVIQSTAVTTSVSGTKEQIKG 116  
DB 96 RIKRKRELTIOKARETVQKNTSSNYIDAVINADSL-ADAVRIQAMSTIVKANQDILVQ-- 152  
QY 117 TGDEKKAKEIKKEKKEKKKQOOSIA 142  
DB 153 --QKEDKQAVEAKKAKENAKOKELA 176  
RESULT 3  
ID 058907 PRELIMINARY; PRT; 270 AA.  
AC 058907;  
DT 01-AUG-1998 (TREMBlrel. 07, Created)  
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
DE Hypothetical protein PH1189.  
GN PH1189.  
OS Pyrococcus horikoshii.  
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
OC Pyrococcus.  
RN NCBI\_TaxID=53953;  
[1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=OT3;  
MEDLINE=98344137; PubMed=9679194;  
RA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,  
RA Yamamoto S., Sekine M., Baba S.-I., Kousugi H., Hosoyama A., Nagai Y.,  
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,  
RA Funahashi T., Tanaka T., Kudo Y., Yamazaki J., Kushiida N., Oguchi A.,  
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,  
RA Masuchi Y., Shizuya H., Kikuchi H.;  
RT "Complete sequence and gene organization of the genome of a hyper-  
thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";  
RL DNA Res. 5:55-76(1998).  
DR EMBL: AP000005; BAA30289.1; -  
KM Hypothetical protein; Complete proteome.  
SQ SEQUENCE 270 AA; 31019 MW; AA54AC8B94FB1EFE CRC64;  
Query Match 17.6%; Score 124.5; DB 17; Length 270;

Best Local Similarity 25.4%; Pred. No. 0.43;  
Matches 44; Conservative 34; Mismatches 44; Indels 51; Gaps 6;  
QY 3 NDAVILRLQK-----GAEDQIIEYLKQVSLKKEKALIQATLREKKLRVENAKL 55  
DB 86 NYVALISSEKPEFVKFSENESTQKLTLEKENEKELKQ--IQMLTKENEDLKEMADL 143  
QY 56 KKEIEELKQELIQAEIQNGVQIAFPSSGTPHLANSWSENVIQSTAVTTSVSGTKEQIKG 115  
DB 144 KRISDLERKHLKAKAQD-----ISLQVQINNLTEENELKEKIAN 185  
QY 116 GTG--DEKKAKE-----IEKKERKEK-----KQOSIACS 144  
DB 186 GTNTINQLAKAKAFLEQGNNEYRTLITKLREQAKSEQSYIEKAKREKLIGS 238  
RESULT 4  
ID Q9V504 PRELIMINARY; PRT; 323 AA.  
AC Q9V504;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
DE CG8235 protein.  
GN CG8235.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;  
OC Ephydroidae; Drosophilidae; Drosophila.  
RN NCBI\_TaxID=7227;  
[1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkley;  
MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter B.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abirl J.F., Agbayani A., An H.-J., Andrews-Plamkosh C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhargava D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,  
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.O., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jatali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Laske P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Modarres C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,  
RA Palazzone M., Pitman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svitzkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Weissman D.A., Weinstock G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195 (2000).  
RN [2]

QY	DB	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20	Q21	Q22	Q23	Q24	Q25	Q26	Q27	Q28	Q29	Q30	Q31	Q32	Q33	Q34	Q35	Q36	Q37	Q38	Q39	Q40	Q41	Q42	Q43	Q44	Q45	Q46	Q47	Q48	Q49	Q50	Q51	Q52	Q53	Q54	Q55	Q56	Q57	Q58	Q59	Q60	Q61	Q62	Q63	Q64	Q65	Q66	Q67	Q68	Q69	Q70	Q71	Q72	Q73	Q74	Q75	Q76	Q77	Q78	Q79	Q80	Q81	Q82	Q83	Q84	Q85	Q86	Q87	Q88	Q89	Q90	Q91	Q92	Q93	Q94	Q95	Q96	Q97	Q98	Q99	Q100
QY	DB	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20	Q21	Q22	Q23	Q24	Q25	Q26	Q27	Q28	Q29	Q30	Q31	Q32	Q33	Q34	Q35	Q36	Q37	Q38	Q39	Q40	Q41	Q42	Q43	Q44	Q45	Q46	Q47	Q48	Q49	Q50	Q51	Q52	Q53	Q54	Q55	Q56	Q57	Q58	Q59	Q60	Q61	Q62	Q63	Q64	Q65	Q66	Q67	Q68	Q69	Q70	Q71	Q72	Q73	Q74	Q75	Q76	Q77	Q78	Q79	Q80	Q81	Q82	Q83	Q84	Q85	Q86	Q87	Q88	Q89	Q90	Q91	Q92	Q93	Q94	Q95	Q96	Q97	Q98	Q99	Q100
QY	DB	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20	Q21	Q22	Q23	Q24	Q25	Q26	Q27	Q28	Q29	Q30	Q31	Q32	Q33	Q34	Q35	Q36	Q37	Q38	Q39	Q40	Q41	Q42	Q43	Q44	Q45	Q46	Q47	Q48	Q49	Q50	Q51	Q52	Q53	Q54	Q55	Q56	Q57	Q58	Q59	Q60	Q61	Q62	Q63	Q64	Q65	Q66	Q67	Q68	Q69	Q70	Q71	Q72	Q73	Q74	Q75	Q76	Q77	Q78	Q79	Q80	Q81	Q82	Q83	Q84	Q85	Q86	Q87	Q88	Q89	Q90	Q91	Q92	Q93	Q94	Q95	Q96	Q97	Q98	Q99	Q100
QY	DB	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20	Q21	Q22	Q23	Q24	Q25	Q26	Q27	Q28	Q29	Q30	Q31	Q32	Q33	Q34	Q35	Q36	Q37	Q38	Q39	Q40	Q41	Q42	Q43	Q44	Q45	Q46	Q47	Q48	Q49	Q50	Q51	Q52	Q53	Q54	Q55	Q56	Q57	Q58	Q59	Q60	Q61	Q62	Q63	Q64	Q65	Q66	Q67	Q68	Q69	Q70	Q71	Q72	Q73	Q74	Q75	Q76	Q77	Q78	Q79	Q80	Q81	Q82	Q83	Q84	Q85	Q86	Q87	Q88	Q89	Q90	Q91	Q92	Q93	Q94	Q95	Q96	Q97	Q98	Q99	Q100
QY	DB	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20	Q21	Q22	Q23	Q24	Q25	Q26	Q27	Q28	Q29	Q30	Q31	Q32	Q33	Q34	Q35	Q36	Q37	Q38	Q39	Q40	Q41	Q42	Q43	Q44	Q45	Q46	Q47	Q48	Q49	Q50	Q51	Q52	Q53	Q54	Q55	Q56	Q57	Q58	Q59	Q60	Q61	Q62	Q63	Q64	Q65	Q66	Q67	Q68	Q69	Q70	Q71	Q72	Q73	Q74	Q75	Q76	Q77	Q78	Q79	Q80	Q81	Q82	Q83	Q84	Q85	Q86	Q87	Q88	Q89	Q90	Q91	Q92	Q93	Q94	Q95	Q96	Q97	Q98	Q99	Q100
QY	DB	Q1																																																																																																			

Dd    84 / N-EKSEVKNVEAKNEVKSEVSKSVNKA-----NEKSEVKS SVKS 88

Qy    108 GTEQIKGCTDEBKAKEKI EKKGEKKCKO 138  
       |::| :|| :: ||| ||| ::  
Db    888 EVKSEVSEVNKVSEVSEENKAEKAR 916

RESULT 7	
014527	
ID 014527	PRELIMINARY;
	PRT; 1024 AA.

DT 01-JAN-1998 (TREMblrel. 05, Created)  
DT 01-OCT-2001 (TREMblrel. 18, last sequence update)  
DT 01-OCT-2002 (TREMblrel. 22, last annotation update)  
DE Hypothetical protein KIAA0291 (Fragment).

OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;

RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RA Ohara O., Nagase T., Ishikawa K., Nakajima D., Chira M., Seki N.,  
RA Nomura N.;

RL Submitted (Aug-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB006629; BAA22960.2; -.  
 DR InterPro; IPR000938; CAP-Gly.  
 DR Pfam; PF01302; CAP\_GLY; 2.

DR	PROSITE; PS50245; CAP_GLY_2; 2.
KM	Hypothetical protein.
FT	NON_TER 1 1
SQ	SEQUENCE 1024 AA; 112981 MW;

Query Match	15.7%	Score 11.5;	DB 4;	Length 1024;
Best Local Similarity	23.0%;	Pred. No. 13;		
Matches	38;	Conservative	43;	Mismatches 49;
			Indels	35;
			Gaps	5

```

Dy      68 QAEIQ-NGVKQIAFPBPGSTPLHANSMTSENVIOS-----TAVTVSSG 106
Db      713 LEKLDVYRGQAQAIIEFLKEQISLAEKWL-----DYERLQRAEAKQKEVESLREKL 766

```

```

Db 767 VAENRLOAVALSSQHTHMIESNDISEETITKEIVBSLQJKNKRREYVALTISQTEM 8228
Qy 109 TKEDL-----KGGTDEK-----KAKEKTEKKGKEXKQKQSIAQ 143
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 8	
Q9UDT6	
ID Q9UDT6	PRELIMINARY;
	PRT; 1046 AA.

DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)  
DE Hypothetical protein KIAA0291.

OS Homo sapiens (Human) ,  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;

RP SEQUENCE FROM N.A.  
RX MEDLINE=99063792; PubMed=9847074;  
RA Sulston J.E., Waterston R.;  
RT "Toward a complete human genome s

RN	[2]	acc.	1400 (1.220)
22		0.407	

RP SEQUENCE FROM N.A.  
RA Bemis G., Langston Y., Kellen J.;  
RT "The sequence of Homo sapiens BAC clone GS1-166C5."  
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases  
RM [3]

RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases  
DR EMBL; AC005015; AAF03524.1; -  
DR Genew; HGNC:2586; CYLN2.

DR Pfam; PF01302; CAP\_GLY\_2.  
DR PROSITE; PS00845; CAP\_GLY\_1; 2.  
DR PROSITE; PS0245; CAP\_GLY\_2; 2.  
KM Hypothetical protein.

Query Match	15.7%	Score 111.5;	DB 4;	Length 1046;
Best Local Similarity	23.0%	Pred. No. 13;		
Matches 38;	Conservative 43;	Mismatches 49;	Indels 35;	Gaps 5

```
07      8 LKLEQKGMAODIIEYLTQQVSLKEKAILOATLREKKIKRVENAKIKEIELKQELI   67
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     735 LEKLIDVEYNGQAQIAIFLEKQISLAETGVL-----DYERLQRAEAQCKQEVESTREKL  788
```

Db	789	VAENRLQAVEALCSSQHTMIESNDISEBTIRKTEYEGLODKLNKDKVEYALTQSOTEM	848
Qy	109	TKHQI-----KGTGDEK-----YAKEKIEKKGEKKKKQOOSING	143

Db 849 LRAVSHLESCKSGEKNDALLKERRELEETVSRKTHDASG 892

ID	Q8T5C7	PRELIMINARY;	PRT; 2055 AA
AC	Q8T5C7;		
DT	01-JUN-2002	(TREMBLrel. 21, Created)	
DT	01-JUN-2002	(TREMBLrel. 21, Last sequence up	

DE Chimeric erythrocyte-binding protein MAEBL.  
GN MAEBL.  
OS Plasmodium falciparum.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; P

RN [1] -  
RP SEQUENCE FROM N.A.  
RA Blair B.L., Kappe S.H.I., Maciel J.E., Balu D.J., Adams B.:  
RT "Plasmodium falciparum maeli is a unique member of the ebl family."

RA Michon P., Stevens J.R., Kaneko O., Adams J.H.;  
RT "Plasmodium falciparum maebl is a unique member of the EBL family.";  
RL Mol. Biol. Evol. 0:0-0(2002).

RP SEQUENCE FROM N.A.  
RA Ghai M., Dutta S., Ockenhouse C.F.;  
RT "Identification, expression and functional characterization of MAEBL  
a sporozoite and asexual blood stage chimeric erythrocyte binding

RL	Mol. Biochem. Parasitol.	0:0-0(2002).
DR	EMBL; AY042084; AAL10509.1;	-.
DR	EMBL; AF400002; AAM90625.1;	-.
DR	HSP; P04268; 1IC2.	

Query Match	15.7%	Score 111.5;	DB 5;	Length 2055;
Best Local Similarity	28.7%	Pred. No. 27;		
Matches 39;	Conservative 24;	Mismatches 40;	Indels 33;	Gaps 3

QY 4 NDAVLKRLKQGAADQIIEYLKQVSLKEKAILQATLREEKLRVENAKLKEIEELK 63



**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2004, 18:18:32 ; Search time 16.1412 Seconds  
(without alignments)  
875.822 Million cell updates/sec

Title: US-09-930-169-1

Perfect score: 709  
Sequence: 1 MANNDAVLKRLBQKGAADQ.....EKKEKKEKQSIAGSADS 147

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: PIR\_76:.\*  
2: PIR1:.\*  
3: PIR2:.\*  
4: PIR3:.\*  
5: PIR4:.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	682	96.8	310	2 B55053	endothelial monocy
2	509	71.6	310	2 B55053	endothelial monocy
3	125.5	17.7	507	2 S05342	hypothetical prote
4	124.5	17.6	670	2 G71061	hypothetical prote
5	117	16.5	262	2 D54078	methy-accepting c
6	116.5	16.4	1079	2 T38913	translational initia
7	110	15.5	947	2 T20156	hypothetical prote
8	109.5	15.4	244	2 T26265	hypothetical prote
9	108.5	15.3	662	2 A54078	methy-accepting c
10	108	15.2	456	2 B86903	hypothetical prote
11	108	15.2	461	2 JN0097	secreted 45k prote
12	108	15.2	1087	2 T30330	gelsoin-related p
13	107	15.1	1690	2 T13030	microtubule bindin
14	106.5	15.0	1093	2 A47212	transcription fact
15	104	14.7	473	2 F70031	cell wall-binding
16	104	14.7	516	2 D64410	replication factor
17	104	14.7	1526	2 A45605	mature-parasite-in
18	103.5	14.6	561	2 B84050	hypothetical prote
19	103	14.5	286	2 B64536	hypothetical prote
20	103	14.5	1130	2 T34081	hypothetical prote
21	103	14.5	1169	2 A64505	plis homolog - Met
22	102.5	14.5	853	2 T51505	hypothetical prote
23	102	14.4	405	2 S42875	diacylglycerol S
24	102	14.4	530	2 B82189	methy-accepting c
25	102	14.4	1132	2 T43483	translational initia
26	101.5	14.3	433	2 T71853	hypothetical prote
27	101.5	14.3	1046	2 T42720	cytoplasmic linker
28	101	14.2	1365	2 T45031	hypothetical prote
29	101	14.2	1938	1 JX0178	myosin heavy chain

30	100	14.1	777	2 T21048	hypothetical prote
31	99.5	14.0	886	2 H69378	conserved hypothet
32	99.5	14.0	2116	2 A26655	myosin heavy chain
33	99	14.0	392	2 G95258	secreted 45 kd pro
34	99	14.0	392	2 B98124	general stress pro
35	99	14.0	1388	2 T30335	KUP2 protein - Afr
36	99	14.0	1940	2 A29320	myosin heavy chain
37	98.5	13.9	407	2 F70461	hypothetical prote
38	98.5	13.9	1046	2 T42734	cytoplasmic linker
39	98.5	13.9	1095	2 F96712	hypothetical prote
40	98.5	13.9	1115	2 B84476	probable TPR repa
41	98.5	13.9	1179	2 F71190	probable chromosom
42	98.5	13.9	1827	2 T16270	hypothetical prote
43	98	13.8	1269	2 F84730	probable myosin he
44	97.5	13.8	553	2 E91120	probable membrane
45	97.5	13.8	553	2 B85965	probable membrane

## ALIGNMENTS

RESULT 1  
B55053  
endothelial monocyte-activating protein II precursor - human  
C/Species: Homo sapiens (man)  
C/Date: 08-Jul-1995 #sequence\_revision 03-Aug-1995 #text\_change 09-Dec-2002  
C/Accession: B55053  
R/Kao, J.; Houck, K.; Fan, Y.; Haehnel, I.; Libutti, S.K.; Kayton, M.L.; Grikscheit, T.;  
J.; Stern, D.M.  
J. Biol. Chem. 269, 25106-25119, 1994  
A/Title: Characterization of a novel tumor-derived cytokine. Endothelial-monocyte activa  
A/Reference number: A55053; MUID:95014290; PMID:7929199  
A/Accession: B55053  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-310 <KAO>  
A/Cross-references: GB:U10117  
C/Superfamily: endothelial monocyte-activating polypeptide II precursor (pro-EMAP II)

Query Match 96.2%; Score 682; DB 2; Length 310;  
Best Local Similarity 98.0%; Pred. No. 1.5e-37;  
Matches 144; Conservative 0; Mismatches 1; Indels 2; Gaps 1;

Qy	1	MANNDAVLKRLBQKGAADQIEYLKQVSLKELKATLQATLEKKLRVENAKLKEKE	60
Db	1	MANNDAVLKRLBQKGAADQIEYLKQVSLKELKATLQATLEKKLRVENAKLKEKE	60

Qy	61	ELKQELIQAEIQNGVQKIRFSGTPLHNSMVSENVISTAVTTVSSGTGEQIKGTGDE	120
Db	61	ELKQELIQAEIQNGVQKIRFSGTPLHNSMVSENVISTAVTTVSSGTGEQIKG--GDE	118

Qy	121	KKAKETIEKKGEKKKKQSIAGSADS	147
Db	119	KKAKETIEKKGEKKKKQSIAGSADS	145

RESULT 2  
A55053  
endothelial monocyte-activating protein II precursor - mouse  
C/Species: Mus musculus (house mouse)  
C/Date: 08-Jul-1995 #sequence\_revision 03-Aug-1995 #text\_change 09-Dec-2002  
C/Accession: A55053; A44032  
R/Kao, J.; Houck, K.; Fan, Y.; Haehnel, I.; Libutti, S.K.; Kayton, M.L.; Grikscheit, T.;  
J.; Stern, D.M.  
J. Biol. Chem. 269, 25106-25119, 1994  
A/Title: Characterization of a novel tumor-derived cytokine. Endothelial-monocyte activa  
A/Reference number: A55053; MUID:95014290; PMID:7929199  
A/Accession: A55053  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-310 <KAO>  
A/Cross-references: GB:U10118; NID:g498911; PIDN:AAA62203.1; PID:g498912  
R/Kao, J.; Ryan, J.; Brett, G.; Chen, J.; Shen, H.; Fan, Y.G.; Godman, G.; Familletti, P

J. Biol. Chem. 267, 20239-20247, 1992  
 A/Title: Endothelial monocyte-activating polypeptide II. A novel tumor-derived polypeptide  
 A/Reference number: A44032; MUID:93015897; PMID:1400342  
 A/Accession: A44032  
 A/Molecule type: Protein  
 A/Residues: 145-158, 'X', 160-164 <KA2>  
 A/Experimental source: methylcholanthrene A fibrosarcoma cells  
 A/Note: sequence extracted from NCBI backbone (NCBIP:115676)  
 C/Superfamily: endothelial monocyte-activating polypeptide II precursor (pro-EMAP II)

Query Match 71.8%; Score 509; DB 2; Length 310;  
 Best Local Similarity 76.4%; Pred. No. 2,7e-26;  
 Matches 113; Conservative 8; Mismatches 23; Indels 4; Gaps 3;

QY 1 MANDAVLKRLBQKGAADQIIIEYLKQVSLKKEKAILQATLREKKLRVENAKLKEIE 60  
 DB 1 MANDAVLKRLBQKGAADQIIIEYLKQVSLKKEKAILQATLREKKLRVENAKLKEIE 60  
 QY 61 ELKQELIQAEIQNGVQKQIAPPSGTPPLHANSWSENVQSTAV-TTVSSGTEQIKGCTGD 119  
 DB 61 ELKQELIQAEIQNGVQKQIAPPSGTPPLHANSWSENVQSTAV-TTVSSGTEQIKGCTGD 119  
 QY 120 EKPAKEIKKKEKKEKKQKQSIAGSADS 147  
 DB 119 EKPAKEIKKKEKKEKKQKQSIAGSADS 145

RESULT 3  
 S05542  
 hypothetical protein, 54K - Enterococcus faecium  
 C/Species: Enterococcus faecium  
 C/Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 15-Oct-1999  
 C/Accession: S05542  
 R/Fuerset, P.; Moeser, H.U.; Solter, M.  
 Nucleic Acids Res. 17, 6724, 1989  
 A/Title: A protein of unusual composition from Enterococcus faecium.  
 A/Reference number: S05542; MUID:89385998; PMID:2780297  
 A/Accession: S05542  
 A/Molecule type: DNA  
 A/Residues: 1507 <FUE>  
 A/Cross-references: GB:X16421, EMBL:M26048, NID:g43333, PIDN:CAA34442.1; PID:g43334  
 A/Note: the authors translated the codon CGT for residues 221 and 223 as Lys

Query Match 17.7%; Score 125.5; DB 2; Length 507;  
 Best Local Similarity 28.1%; Pred. No. 0.38;  
 Matches 41; Conservative 33; Mismatches 51; Indels 15; Gaps 5;

QY 5 DAVLKRLBQKGAADQIIIEYLKQVSLKKEKAILQATLREKKLRVENAKLKEIEELKO 64  
 DB 29 DKTIADLQNGQASQSGIEALBEGVSAINTKA--QDLTLKQDTLRKESAGLQKQEIKDLOE 86  
 QY 65 --ELIQAEIQNGVQKQIAPPSGTP-----LHANSWSENVQSTAVTTVSSGTEQIKG 116  
 DB 87 RIKRREATTQKQARETQVKNKTSNYIDAVINADSL-ADAVGRIQAMSTIVKANQDQIVQ-- 143  
 QY 117 TGDEKAKKEIKKKEKKKQKQSI 142  
 DB 144 --QKEDKQAVEAKKAKENAKKELA 167

RESULT 4  
 G71061  
 hypothetical protein PH189 - Pyrococcus horikoshii  
 C/Species: Pyrococcus horikoshii  
 C/Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 05-Nov-1999  
 C/Accession: G71061  
 R/Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin  
 M.; Ohfuku, Y.; Punahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi  
 DNA Res. 5, 55-76, 1998  
 A/Title: Complete sequence and gene organization of the genome of a hyper-thermophilic  
 A/Reference number: A71000; MUID:98344137; PMID:9679194  
 A/Accession: G71061  
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA  
 A/Residues: 1-270 <KAW>  
 A/Cross-references: GB:AP000005; NID:g3236132; PIDN:BA30289.1; PID:d1031232; PID:g32576  
 A/Experimental source: strain O73  
 A/Note: this accession replaces an interim accession for a sequence replaced by GenBank  
 C/Genetics:  
 A/Genes: PH1189

Query Match 17.6%; Score 124.5; DB 2; Length 270;  
 Best Local Similarity 25.4%; Pred. No. 0.23;  
 Matches 44; Conservative 34; Mismatches 44; Indels 51; Gaps 6;

QY 3 MNDVAVLKRLBQK-----GAEADQIIIEYLKQVSLKKEKAILQATLREKKLRVENAKL 55  
 DB 86 NYVALISISEKPEYKFSNSETSQLETLKENEKELKQ--IQULTKENEQKKNENAO 143  
 QY 56 KEIEELKQELIQAEIQNGVQKQIAPPSGTPPLHANSWSENVQSTAVTTVSSGTEQIKG 115  
 DB 144 KRISDLEHGLKEAKAQD-----ISELQVQINNLTKENELKEKIAN 185  
 QY 116 GTG--DEKAKKK-----IEKKGEKKK-----KQOSIAGS 144  
 DB 186 QNTTITLQKAKKFLBQNNERTLITLLEBQAKKSEOSTYKAKRKLIGS 238

RESULT 5  
 D54078  
 methyl-accepting chemotaxis protein tlpB - Bacillus subtilis  
 N/Alternate names: Chemotaxis transducer homolog TLPB  
 C/Species: Bacillus subtilis  
 C/Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 20-Jun-2000  
 C/Accession: D54078; A69724  
 R/Hanlon, D.W.; Ordal, G.W.  
 J. Biol. Chem. 269, 14038-14046, 1994  
 A/Title: Cloning and characterization of genes encoding methyl-accepting chemotaxis prote  
 A/Reference number: A54078; MUID:94245722; PMID:8188684  
 A/Accession: D54078  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-662 <HAN>  
 A/Cross-references: GB:L29189; NID:g459687; PIDN:AAA20557.1; PID:g459691  
 A/Note: authors translated the codon CAG for residue 10 as Trp  
 R/Kunze, F.; Ogasawara, N.; Moser, I.; Albertin, A.M.; Alloni, G.; Azevedo, V.; Bertec  
 C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuno, V.; Carter, N.M.; Cho  
 A.; Enrich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabre, C.; Ferrari, E.  
 Nature 390, 249-256, 1997  
 A/Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, S.; Galizzi, A.; Gallerc  
 lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holzapfel, S.; Hosono, S.; Hullo, M.F.;  
 Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,  
 A/Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Maude, S.; Maueel,  
 Y.M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Paro, V.; Pohl, T.M.; Portetelle,  
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon,  
 A/Authors: Schleich, S.; Schroeter, R.; Scofield, F.; Sekiguchi, J.; Sekowaka, A.; Setor,  
 akuchin, M.; Tanakoshi, A.; Tanaka, T.; Tepstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,  
 T.; Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasunoto, K.; Yata, K.; Yoshida, K.  
 A/Authors: Yoshikawa, H.F.; Zumsattel, B.; Yoshikawa, H.; Danchin, A.  
 A/Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
 A/Reference number: A69580; MUID:98044033; PMID:9384377  
 A/Accession: A69724  
 A/Molecule type: DNA  
 A/Status: nucleic acid sequence not shown; translation not shown  
 A/Residues: 1-662 <KUN>  
 A/Cross-references: GB:Z99119, GB:Z99120; GB:AL009126; NID:g2655613; PIDN:CAB15112.1; PII  
 A/Experimental source: strain 168  
 C/Genetics:  
 A/Genes: tlpB  
 C/Superfamily: probable methyl-accepting chemotaxis transducer  
 C/Keywords: transmembrane protein

Query Match 16.5%; Score 117; DB 2; Length 662;  
 Best Local Similarity 26.0%; Pred. No. 1.8;  
 Matches 45; Conservative 26; Mismatches 64; Indels 38; Gaps 5;

OY 5 DAVLKRLKLEQKGAADQIIEYIK--QQVSLKEKAILQA-----TLREKKLR 49  
Db 463 EAVVKGLETQKQDITSLINVLNIGADQTNLALNAITEAARAGEYGRGFSVAABEVKRLA 522  
OY 50 VENAKLKEIEELKQELIQ-----AEIONGKQIAPFSGTPLHANSWSENVIOST 100  
Db 523 VQSDAKSEIKELGLIOEIVREISTSLSMFQSVNHEVKE-----GLQITDQTAFSEKQIY 575  
OY 101 AVTTVSSG-----TKEQIKGTGDEKKAKIEKKEKKEKQOOSIAGSAD 146  
Db 576 EMTTQISELQNMNATVQIAGSGQEVSSAVEDISAIVAKESAGQIDIAAE 628

## RESULT 6

T28913  
Translation initiation factor if-2 - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 02-Sep-2000  
C:Accession: T38913  
R:Pearson, D.; Churchev, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.  
submitted to the EMBL Data Library, February 1996  
A:Reference number: Z21817  
A:Accession: T38913  
A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA  
A:Residues: 1-1079 <PEA>  
A:Cross-references: EMBL:Z69728; PIDN:CAA93574.1; GSPDB:GN00066; SPDB:SPAC56F8.03  
A:Experimental source: strain 972h-, cosmid c56F8  
C:Genetics:  
A:Gene: SPDB:SPAC56F8.03  
A:Map position: 1  
A:Superfamily: translation elongation factor Tu homology  
C:Superfamily: translation elongation factor Tu homology <FTU>  
F:485-612/Domain: translation elongation factor Tu homology

Query Match 16.4%; Score 116.5; DB 2; Length 1079;  
Best Local Similarity 21.6%; Pred. No. 3.3;  
Matches 38; Conservative 41; Mismatches 58; Indels 39; Gaps 5;

OY 3 NNDVAVLKRLKLEQKGAADQIIEYIKQVSLKEKAIL-----QATLREKKLRVNA 53  
Db 246 NVTALQKMLEKRRAREEEOIRREBARIABEERLAVBEARKEEALKKKEKKEKKE 305  
OY 54 KLR-----KEIEELKQELIQAEIONGKQIAPFSG-----TPLHAN----- 89  
Db 306 EMKQKGYKLSKKQKQQLAQRLOQMLESQVAVGSLNGEKKQKPYVTNNKKSRSSTS 365  
OY 90 SMSENVIOSTAVTTVSSGTEQIKGTGDEKKAKIEKKEKKEKQOOSIAGSA 145  
Db 366 SISSGILSESPATISIVDEPOK-----DSKDSSEKVEKETEVERKEENEAEAA 415

## RESULT 7

T20156  
hypochemical protein F57F5.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T20156; T22857  
R:Holt, R.

submitted to the EMBL Data Library, August 1996  
A:Reference number: Z19230  
A:Accession: T20156

A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-947 <WII>

A:Cross-references: EMBL:Z78012; PIDN:CAB01418.1; GSPDB:GN00023; CESP:F57F5.2  
A:Experimental source: clone C52E4  
R:Harris, B.

submitted to the EMBL Data Library, July 1996  
A:Reference number: Z19627  
A:Accession: T22857

A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-947 <W12>

A:Cross-references: EMBL:Z75953; PIDN:CAB00103.1; GSPDB:GN00023; CESP:F57F5.2  
A:Experimental source: clone F57F5  
C:Genetics:  
A:Gene: CESP:F57F5.2  
A:Map position: 5  
A:Introns: 24/3; 192/2; 313/1; 466/2; 574/3; 665/2; 705/2; 738/3; 832/3; 876/3; 906/3

Query Match 15.5%; Score 110; DB 2; Length 947;  
Best Local Similarity 21.6%; Pred. No. 7.6;  
Matches 37; Conservative 43; Mismatches 59; Indels 32; Gaps 6;

OY 1 MANDVAVLKRLKLE---QKGAADQIIEYIKQVSLKEKAILQATLREKKLRVNAKLR 57  
Db 741 LSSSDILAVSVENSKQSDGEPPTSSSELKEVNRIRREALAQEK-EEERTTKEENQKIEE 799  
OY 58 EIEELKQELIQAEIONGKQIAPFSGTPLHANSWSENVIO 98  
Db 800 VGEDEVSEATSLDSEVSHDNNISFQMPDSIPIHEDRTSLPAPTSETIGDAISKKLE 859  
OY 99 ----STAVTTVSSGTEQIKGTGDEKKAKIEKKEKKEKQOOSIAGSA 145  
Db 860 KEDNSMSISLDERTTYSAPFTT-----TRLNQLKLEKKEKRSWAGSS 905

## RESULT 8

T26265  
hypochemical protein W07G1.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T26265  
R:Percy, C.

submitted to the EMBL Data Library, November 1996  
A:Reference number: Z20186  
A:Accession: T26265

A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-244 <W12>

A:Cross-references: EMBL:Z82076; NID:el247330; PIDN:CAB04933.1; GSPDB:GN00020; CESP:W07G  
A:Experimental source: clone W07G1  
C:Genetics:  
A:Gene: CESP:W07G1.1  
A:Map position: 2  
A:Introns: 49/3; 161/1; 210/3

Query Match 15.4%; Score 109.5; DB 2; Length 244;  
Best Local Similarity 24.8%; Pred. No. 1.9;  
Matches 37; Conservative 35; Mismatches 54; Indels 23; Gaps 5;

OY 10 RLEQKGAADQIIEYIKQVSL-LKEKAILQATLREKKLRVENAKLKEIEE 61  
Db 75 KLRKKKKRSGQVQPKTEHIELGDEISIPLNEVVLKKKKKKKLVQKNLDLHGSIF 134  
OY 62 LKQEL---IQAEIONGKQIAPF---SGTPLHANSWSENVIOSTAVTTVSSGTEQIKG 115  
Db 135 LAETAEIELSEIEDSEDEHEHRRSATPDHRRPATKSHVAPSRASSR-----RG 185  
OY 116 CTGDEKKAKIEKKEKKEKQOOSIAGS 144  
Db 186 SRDEERREDEERKKRDEKRSVSS 214

## RESULT 9

A54078  
methyl-accepting chemotaxis protein mcpB - Bacillus subtilis

C:Species: Bacillus subtilis  
C:Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 20-Jun-2000  
C:Accession: A54078; H69655  
R:Hanlon, D.W.; Ordal, G.W.

J. Biol. Chem. 269, 14038-14046, 1994  
A:Title: Cloning and characterization of genes encoding methyl-accepting chemotaxis prot.  
A:Reference number: A54078; MUID:94245722; PMID:8188684  
A:Accession: A54078  
A:Status: preliminary

A:Molecule type: DNA  
A:Residues: 1-662 <HAN>  
A:Cross-references: GB:U29189; NID:9459687  
A:Note: the sequence in Genbank entry BACMCPTR, release 106.0, (PID:9459688) has 360-N  
R:Kunze, P.; Ogasawara, N.; Moszer, I.; Alibertini, A.M.; Alloni, G.; Azevedo, V.; Berter  
C.; Bron, S.; Brouillet, S.; Bueschi, C.V.; Caldwell, V.; Capuano, V.; Carter, N.M.; Cio  
A.; Ehrlich, S.D.; Emmerson, P.T.; Eutian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
Nature 390, 249-256, 1997  
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galluzzi, A.; Gall  
ieb, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.P.  
Koetter, P.; Koningsveld, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,  
A.; Luthers, J.; Lazarevic, V.; Lee, S.W.; Levine, A.; Liu, H.; Maeda, S.; Maue  
Y., M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portebell  
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon,  
A.; Schuster, S.; Schuster, R.; Scott, P.; Seifried, J.; Sekowski, A.; Serd  
T.; Winters, P.; Wipit, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K  
A.; Yoshida, Y.; Yoshida, H.; Yoshida, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.  
A:Reference number: A69580; MUID:98044033; PMID:9384377  
A:Accession: H69655  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-359, 'N', 361-449, 'G', 451-662 <RUN>  
A:Cross-references: GB:299119; GB:299120; GB:AL009126; NID:92635613; PID:NCAB15115.1; PI  
A:Experimental source: strain 168  
C:Comment: Strains in which this protein is inactivated are defective in chemotaxis tow  
C:Genetics:  
A:Gene: mcpB  
C:Superfamily: probable methyl-accepting chemotaxis transducer  
C:Keywords: chemotaxis; signal transduction; transmembrane protein

Query Match 15.3%; Score 108.5; DB 2; Length 662;  
Best Local Similarity 26.8%; Pred. No. 6.5; Indels 37; Gaps 7;  
Matches 44; Conservative 27; Mismatches 56; Indels 37; Gaps 7;

QY 5 DAVLKRLKQKGAADQIIEYLK--QGVSLKEKAILQA-----TLREKKLR 49  
DB 464 EAVVKGEGSKSIOITSLRVINGIADQTNLALNAEAPAGSGRGFSVAEVRKLA 533  
QY 50 VENAALKEIEELKQELIQAEIONGVQIAFPSTGPLHANSWVSNVIOSTAVTTVSSGT 109  
DB 524 VQASADSAKEIEKIQEIV- AEID-----TSLHMFKEVQEV-QSGLV--VTDNT 568

QY 110 KEQIKG-----GTDEKAKKEIEKKEKKEKQKQSIAGSD 146  
DB 569 KESFQSTFMTNTEIAGKLQTMNSTVEQLSPRSQVSAVSGIAD 612

RESULT 10  
E86903  
hypothetical protein usp45 [imported] - Lactococcus lactis subsp. lactis (strain IL1403)  
C:Species: Lactococcus lactis subsp. lactis  
C>Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 03-Aug-2001  
C:Accession: E86903  
R:Bojorin, A.; Wincker, P.; Manger, S.; Jallion, O.; Malarme, K.; Weissenbach, J.; Ehrli  
Genome Res. 11, 731-753, 2001  
A:Title: The complete genome sequence of the lactic acid bacterium *Lactococcus lactis* s  
A:Reference number: A86625; MUID:21235186; PMID:11337471  
A:Molecule type: DNA  
A:Residues: 1-456 <STO>  
A:Cross-references: GB:AE005176; PID:g12725296; PIDN:AAK6327.1; GSPDB:GN00146  
A:Experimental source: strain IL1403  
C:Genetics:  
A:Gene: usp45

Query Match 15.2%; Score 108; DB 2; Length 456;  
Best Local Similarity 26.6%; Pred. No. 4.7; Indels 65; Gaps 6;  
Matches 42; Conservative 33; Mismatches 65; Indels 18; Gaps 6;

QY 1 MANNDVLRLEQKGAADQIIEYLKQVSLKEKAILQATLR-----EKKLRVEN 52

DB 33 IAKDATTSSAQAQAQVDSLSQKVSQKQK---QASTKAQIAKIESAKKLNAGI 89  
QY 53 AKLKEIEELKQELIQAEIONGVQIAFPSTGPLHANSWVSNVIOSTAVTTVSSGTJE 111  
DB 90 ATLNEISIAE-RKTLLEAQAARSQAQVNSATNVMDAVNSSLTDVIOKTAITVSSANKQ 148

QY 112 QIKGTGDEKAKKEIEKKEKKEKQK--SIAGSADS 147  
DB 149 ME---QOEKQKELSQSEYTKVKNVQVSLQSIDS 183

RESULT 11  
JN0097  
secreted 45K protein precursor - Lactococcus lactis  
C:Species: Lactococcus lactis  
C>Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 16-Jul-1999  
C:Accession: JN0097  
R:van AseelDonk, M.; Rutten, G.; Oteman, M.; Slezzen, R.J.; de Vos, W.M.; Simons, G.  
Gene 95, 155-160, 1990  
A:Title: Cloning of usp45, a gene encoding a secreted protein from *Lactococcus lactis* aut  
A:Reference number: JN0097; MUID:91071599; PMID:2123812  
A:Accession: JN0097  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-461 <VAN>  
A:Cross-references: GB:M35374

Query Match 15.2%; Score 108; DB 2; Length 461;  
Best Local Similarity 26.5%; Pred. No. 4.8; Indels 12; Gaps 5;  
Matches 41; Conservative 33; Mismatches 69; Indels 12; Gaps 5;

QY 1 MANNDVLRLEQKGAADQIIEYLKQVSLKE-----KAILQATLRREKKLRVENAKL 55  
DB 33 IAKDATTSSAQAQAQVDSLSQKVSQKQKOTSKAQIAKIESAKKLNAGIATL 92

QY 56 KKEIEELKQELIQAEIONGVQIAFPSTGPLHANSWVSNVIOSTAVTTVSSGTJEQIK 114  
DB 93 NSIKR-RKTLLEAQAARSQAQVNSATNVMDAVNSSLTDVIOKTAITVSSANKQMLE 151

QY 115 GGTDEKAKKEIEKKEKKEKQK--SIAGSADS 147  
DB 152 ---QOEKQKELSQSEYTKVKNVQVSLQSIDS 183

RESULT 12  
T30330  
gelsoilin-related protein GRP125 - slime mold (*Dictyostelium discoideum*)  
C:Species: Dictyostelium discoideum  
C>Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999  
C:Accession: T30330  
R:Stocker, S.; Hiery, M.; Marriot, G.  
Mol. Biol. Cell 10, 161-178, 1999  
A:Title: Photoreactive migration of *Dictyostelium* cells is linked to a new type of gelsoilin  
A:Reference number: T20823; MUID:99096692; PMID:9880334  
A:Accession: T30330  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1087 <STO>  
A:Cross-references: EMBL:U95159; NID:g4100185; PID:g4100186; PIDN:AD00774.1  
C:Genetics:  
A:Introns: 137/1

Query Match 15.2%; Score 108; DB 2; Length 1087;  
Best Local Similarity 27.1%; Pred. No. 12; Indels 32; Gaps 6;  
Matches 39; Conservative 28; Mismatches 45; Indels 32; Gaps 6;

QY 19 DQTEYLKQVSLKEKAILQATLRREKKLRVENAKLKEIEELKQELIQAEIO 72  
DB 708 DPLIST-KQKLERLQKEDLEKQKQOQOQOQKNNKI---VEEKSEVKEEDVK 763

QY 73 NGVKQIAFPSTGPLHANSWVSNVIOSTAVTTVSSGTJEQIKGTGD-----EK 121

```
Search completed: January 2, 2004, 18:22:59
Job time : 17.1412 secs
```

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2004, 18:19:07 ; Search time 16.7176 Seconds  
(without alignments)  
372.044 Million cell updates/sec

Title: US-09-930-169-1

Perfect score: 709  
Sequence: 1 MANNDAVLKRLKQKGAADQ.....EKKGKKKKQKQSIAGSADS 147

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:\*  
1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	704	99.3	312	US-08-360-821B-36	Sequence 36, Appli
2	650	91.7	310	US-08-129-456A-37	Sequence 37, Appli
3	509	71.8	310	US-08-129-456A-36	Sequence 36, Appli
4	509	71.8	310	US-08-705-868-3	Sequence 3, Appli
5	509	71.8	310	US-09-123-615-3	Sequence 3, Appli
6	509	71.8	310	US-08-360-821B-35	Sequence 35, Appli
7	109	15.4	461	US-08-186-222-2	Sequence 3, Appli
8	106.5	15.0	414	PCT-US93-03077-3	Sequence 2, Appli
9	106.5	15.0	1093	PCT-US93-03077-1	Sequence 1, Appli
10	104	14.7	961	US-09-914-259-66	Sequence 66, Appli
11	102	14.4	290	US-08-903-801-1	Sequence 1, Appli
12	102	14.4	290	US-09-295-055-1	Sequence 1, Appli
13	102	14.4	431	US-09-286-981B-3	Sequence 3, Appli
14	101.5	14.3	271	US-08-664-596B-11	Sequence 11, Appli
15	101	14.2	956	US-09-914-259-17	Sequence 17, Appli
16	101	14.2	957	US-09-914-259-16	Sequence 16, Appli
17	100.5	14.2	419	US-09-286-981B-15	Sequence 15, Appli
18	100	14.1	588	US-08-714-741-42	Sequence 42, Appli
19	100	14.1	864	US-08-714-741-40	Sequence 40, Appli
20	98.5	13.9	525	US-09-107-532A-5095	Sequence 482, App
21	98	13.8	422	US-09-071-035-484	Sequence 484, App
22	98	13.8	449	US-09-071-035-482	Sequence 3, Appli
23	97	13.7	234	US-08-903-801-3	Sequence 3, Appli
24	97	13.7	234	US-09-295-055-3	Sequence 3, Appli
25	96.5	13.6	210	US-09-222-938A-67	Sequence 67, Appli
26	95	13.4	262	US-09-595-684B-31	Sequence 31, Appli
27	94	13.3	413	US-09-286-981B-5	Sequence 5, Appli

28	94	13.3	667	US-09-071-709-9	Sequence 9, Appli
29	93.5	13.2	459	US-09-071-709-1	Sequence 1, Appli
30	93	13.1	498	US-09-107-532A-6991	Sequence 6991, Ap
31	93	13.1	534	US-09-103-664A-2	Sequence 2, Appli
32	93	13.1	1388	US-08-685-576-1	Sequence 1, Appli
33	92.5	13.0	251	US-09-286-981B-4	Sequence 4, Appli
34	92.5	13.0	708	US-08-235-836C-76	Sequence 76, Appli
35	92.5	13.0	803	US-09-154-750A-85	Sequence 85, Appli
36	92	13.0	414	US-09-286-981B-10	Sequence 10, Appli
37	92	13.0	564	US-09-308-022-6	Sequence 6, Appli
38	92	13.0	631	US-08-847-065-25	Sequence 25, Appli
39	92	13.0	741	US-09-854-856-60	Sequence 60, Appli
40	92	13.0	769	US-09-854-856-44	Sequence 44, Appli
41	92	13.0	801	US-09-854-856-28	Sequence 28, Appli
42	92	13.0	829	US-09-854-856-12	Sequence 12, Appli
43	92	13.0	894	US-09-854-856-54	Sequence 54, Appli
44	92	13.0	922	US-09-854-856-38	Sequence 38, Appli
45	92	13.0	954	US-09-854-856-22	Sequence 22, Appli

#### ALIGNMENTS

RESULT 1  
US-08-360-821B-36  
Sequence 36, Application US/08360821B  
Patent No. 6228837  
GENERAL INFORMATION:  
APPLICANT: Stern, David M.  
APPLICANT: Claus, Matthias  
APPLICANT: Kao, Janet  
APPLICANT: Libutti, Steven K  
TITLE OF INVENTION: Endothelial Monocyte Activating  
TITLE OF INVENTION: Polypeptide II: A Mediator Which Activates Host Response  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham, LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.30, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/360, 821B  
FILING DATE: 08-OCT-96  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 41735-A-PCT-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-278-0400  
TELEFAX: 212-391-0525  
INFORMATION FOR SEQ ID NO: 36:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 312 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-360-821B-36  
Query Match 99.3%; Score 704; DB 3; Length 312;  
Best Local Similarity 99.3%; Pred. No. 1.8e-58;  
Matches 146; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Cy 1 MANNDAVLKRLKQKGAADQIEYLKQVSLKEKAILQATLREKKLRVEMAKKEIE 60

Db 1 MANNDAVLRKLEQKGAADQIIIEYLKQVSLKKEKALQATLREKKLRVENAKLKEIE 60  
QY 61 ELKQELIQAIONGVQIAPPSGTPHLANSMVSENVIOSTAV--TVSSGTEQIKGSGTD 120  
Db 61 ELKQELIQAIONGVQIAPPSGTPHLANSMVSENVIOSTAVTVSSGTEQIKGSGTDE 120  
QY 121 KKAKEKIEKKGEKKKKQOSIAGSADS 147  
Db 121 KKAKEKIEKKGEKKKKQOSIAGSADS 147

RESULT 2  
US-08-129-456A-37  
; Sequence 37, Application US/08129456A  
; Patent No. 5641867  
; GENERAL INFORMATION:  
; APPLICANT: Stern, David M.  
; APPLICANT: Claus, Matthias  
; APPLICANT: Kao, Janet  
; APPLICANT: Kayton, Mark  
; APPLICANT: Libutti, Steven K.  
; TITLE OF INVENTION: Endothelial-Monocyte Activating  
; TITLE OF INVENTION: Polypeptide II: A Mediator which  
; TITLE OF INVENTION: Activates Host Response  
; NUMBER OF SEQUENCES: 37  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Cooper & Dunham LLP  
; STREET: 1185 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0 Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/129,456A  
; FILING DATE: 29-SEP-1993  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: White, John P.  
; REGISTRATION NUMBER: 28,678  
; REFERENCE/DOCKET NUMBER: 41735  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212 278 0400  
; TELEFAX: 212 391 0526  
; INFORMATION FOR SEQ ID NO: 37:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 310 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-129-456A-37

Query Match 91.7%; Score 650; DB 1; Length 310;  
Best Local Similarity 95.9%; Pred. No. 2,1e-53;  
Matches 142; Conservative 0; Mismatches 2; Indels 4; Gaps 3;

QY 1 MANNDAVLRKLEQKGAADQIIIEYLKQVSLKKEKALQATLREKKLRVENAKLKEIE 60  
Db 1 MANNDAVLRKLEQKGAADQIIIEYLKQVSLKKEKALQATLREKKLRVENAKLKEIE 60  
QY 61 ELKQELIQAIONGVQIAPPSGTPHLANSMVSENVIOSTAV--TVSSGTEQIKGSGTD 119  
Db 61 ELKQELIQAIONGVQIAPPSGTPHLANSMVSENVIOSTAVTVSSGTEQIK--AGD 118  
QY 120 EKKAKEIEKKGEKKKKQOSIAGSADS 147  
Db 119 EKKAKEIEKKGEKK--KKOOSIAGSADS 145

RESULT 3  
US-08-129-456A-36  
; Sequence 36, Application US/08129456A  
; Patent No. 5641867  
; GENERAL INFORMATION:  
; APPLICANT: Stern, David M.  
; APPLICANT: Claus, Matthias  
; APPLICANT: Kao, Janet  
; APPLICANT: Kayton, Mark  
; APPLICANT: Libutti, Steven K.  
; TITLE OF INVENTION: Endothelial-Monocyte Activating  
; TITLE OF INVENTION: Polypeptide II: A Mediator which  
; TITLE OF INVENTION: Activates Host Response  
; NUMBER OF SEQUENCES: 37  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Cooper & Dunham LLP  
; STREET: 1185 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0 Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/129,456A  
; FILING DATE: 29-SEP-1993  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: White, John P.  
; REGISTRATION NUMBER: 28,678  
; REFERENCE/DOCKET NUMBER: 41735  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212 278 0400  
; TELEFAX: 212 391 0526  
; INFORMATION FOR SEQ ID NO: 36:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 310 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-129-456A-36

Query Match 71.8%; Score 509; DB 1; Length 310;  
Best Local Similarity 76.4%; Pred. No. 3,5e-40;  
Matches 113; Conservative 8; Mismatches 23; Indels 4; Gaps 3;

QY 1 MANNDAVLRKLEQKGAADQIIIEYLKQVSLKKEKALQATLREKKLRVENAKLKEIE 60  
Db 1 MANNDAVLRKLEQKGAADQIIIEYLKQVSLKKEKALQATLREKKLRVENAKLKEIE 60  
QY 61 ELKQELIQAIONGVQIAPPSGTPHLANSMVSENVIOSTAV--TVSSGTEQIKGSGTD 119  
Db 61 ELKQELIQAIONGVQVAVRSTPLQCTHASEVSQSPVATTSAPATKQIK--AGE 118  
QY 120 EKKAKEIEKKGEKKKKQOSIAGSADS 147  
Db 119 EKKAKEIEKKGEKK--KKOOSIAGSADS 145

RESULT 4  
US-08-705-868-3  
; Sequence 3, Application US/08705868  
; Patent No. 5885798  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga  
; APPLICANT: Coleman, Roger  
; APPLICANT: Au-Young, Janice  
; APPLICANT: Murty, Lynn E.  
; TITLE OF INVENTION: NOVEL MONOCYTE ACTIVATING CYTOKINE



NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: U.S.  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/705,868  
FILING DATE: Filed Herewith  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0117 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 310 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 498910  
US-08-705-868-3

Query Match 71.8%; Score 509; DB 2; Length 310;  
Best Local Similarity 76.4%; Pred. No. 3.5e-40;  
Matches 113; Conservative 8; Mismatches 23; Indels 4; Gaps 3;

QY 1 MANDAVLKRLEQKGAEDQIIIEYIKQOVSLKKEKAIIOATLRREKLRVENAKLKEIE 60  
DB 1 MATNDVAVLKRLEQKGAEDQIIIEYIKQOVALLKKEKAIIOATLRREKLRVENAKLKEIE 60  
QY 61 ELKQELIOAEIIONGVKQIAPSGTPLHANSVSENVIOSTAV-TTVSSGTKEQIKGTGD 119  
DB 61 ELKQELIOAEIIONGVKQIAPSGTPLHANSVSENVIOSTAV-TTVSSGTKEQIKGTGD 119  
QY 120 EKKAKETKEKKEKKKKQSSIASADS 147  
DB 119 EKKVETKEKKEKKKE-KQSSAASTDS 145

RESULT 5  
US-09-123-615-3  
Sequence 3, Application US/09123615  
Patent No. 6090377  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Coleman, Roger  
APPLICANT: Au-Young, Janice  
APPLICANT: Murry, Lynn E.  
TITLE OF INVENTION: NOVEL MONOCYTE ACTIVATING CYTOKINE  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: U.S.  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/123,615  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/705,868  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0117 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 310 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 498910  
US-09-123-615-3

Query Match 71.8%; Score 509; DB 3; Length 310;  
Best Local Similarity 76.4%; Pred. No. 3.5e-40;  
Matches 113; Conservative 8; Mismatches 23; Indels 4; Gaps 3;

QY 1 MANDAVLKRLEQKGAEDQIIIEYIKQOVSLKKEKAIIOATLRREKLRVENAKLKEIE 60  
DB 1 MATNDVAVLKRLEQKGAEDQIIIEYIKQOVALLKKEKAIIOATLRREKLRVENAKLKEIE 60  
QY 61 ELKQELIOAEIIONGVKQIAPSGTPLHANSVSENVIOSTAV-TTVSSGTKEQIKGTGD 119  
DB 61 ELKQELIOAEIIONGVKQIAPSGTPLHANSVSENVIOSTAV-TTVSSGTKEQIKGTGD 119  
QY 120 EKKAKETKEKKEKKKKQSSIASADS 147  
DB 119 EKKVETKEKKEKKKE-KQSSAASTDS 145

RESULT 6  
US-08-360-821B-35  
Sequence 35, Application US/08360821B  
Patent No. 6228837  
GENERAL INFORMATION:  
APPLICANT: Stern, David M.  
APPLICANT: Klaus, Matthias  
APPLICANT: Kao, Janet  
APPLICANT: Kayton, Mark  
APPLICANT: Libutti, Steven K  
TITLE OF INVENTION: Endothelial Monocyte Activating  
TITLE OF INVENTION: Polypeptide II: A Mediator which Activates Host Response  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham, LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.30, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/360,821B  
FILING DATE: 08-OCT-96  
CLASSIFICATION: 435

```

/ ATTORNEY/AGENT INFORMATION:
/ NAME: White, John P.
/ REGISTRATION NUMBER: 28,678
/ REFERENCE/DOCKET NUMBER: 41735-A-PCT-US
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 212-278-0400
/ TELEFAX: 212-391-0525
/ INFORMATION FOR SEQ ID NO: 35:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 310 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: peptide
US-08-360-821B-35

Query Match      71.8%; Score 509; DB 3; Length 310;
Best Local Similarity 76.4%; Pred. No. 3,5e-40;
Matches 113; Conservative 8; Mismatches 23; Indels 4; Gaps 3;

QY 1 MANDAVLRLKRGKGAADQIIIEYLKQVSLKEKALIAQLTLEKKLRVENAKLKEKE 60
DB 1 MANDAVLRLKRGKGAADQIIIEYLKQVSLKEKALIAQLTLEKKLRVENAKLKEKE 60
QY 61 ELKQELIOAEIONGVQIAFPSPGTPPLHANSWSENVIQSTAV-TTVSSGTKEQIKGTGD 119
DB 61 ELKQELIAEIHNGVQVRRLSTPLQTNCTASVSVQSPVATTSATKQIK--AGE 118
QY 120 EKAKERIEKKGKKEKQKQSTINGSADS 147
DB 119 EKKVKEKTEKKGKKE-KQQSAASATDS 145

RESULT 7
US-08-186-222-2
/ Sequence 2, Application US/08186222
/ Patent No. 5559007
/ GENERAL INFORMATION:
/ APPLICANT: Suri, Bruno
/ TITLE OF INVENTION: Bacterial Vectors
/ NUMBER OF SEQUENCES: 4
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: CIBA-GEIGY Corporation
/ STREET: 7 Skyline Drive
/ CITY: Hawthorne
/ STATE: New York
/ COUNTRY: USA
/ ZIP: 10532
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patentin Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/186,222
/ FILING DATE:
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 07/672,205
/ FILING DATE: 19-MAR-1991
/ APPLICATION NUMBER: GB 9006400.7
/ FILING DATE: 22-MAR-1990
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Villamilzar, JoAnn
/ REGISTRATION NUMBER: 30,598
/ REFERENCE/DOCKET NUMBER: 4-17994/A
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (914)785-7121
/ TELEFAX: (914)347-5769
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 461 amino acids

```

```

/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
US-08-186-222-2

Query Match      15.4%; Score 109; DB 1; Length 461;
Best Local Similarity 26.5%; Pred. No. 0.019;
Matches 41; Conservative 33; Mismatches 69; Indels 12; Gaps 5;

QY 1 MANDAVLRLKRGKGAADQIIIEYLKQVSLKEKALIAQLTLEKKLRVENAKL 55
DB 33 IAKQDATTSSAQSAQAQVDSLSQKQSTTKAQIAKIESERKALMQIATL 92
QY 56 KKEIEELKQELIOAEIONGVQIAFPSPGTPPLHANSWSENVIQSTAVTTVSSGTKEQIK 114
DB 93 NESIKE-RFKTLAQARSQVNSATNVDVAVNSKSLTDVQKTAIVTSANQMLE 151
QY 115 GGTDEKAKAKIEKKGKKEKQK--STAGSADS 147
DB 152 ---QQEKEQKELSQKSETVKKNVQFVSLQSLDS 183

RESULT 8
PCT-US93-03077-3
/ Sequence 3, Application PC/TUS9303077
/ GENERAL INFORMATION:
/ APPLICANT: Board of Regents, The University of Texas System
/ APPLICANT: Gaynor, Richard B.
/ APPLICANT: Wu, Poon Kin
/ TITLE OF INVENTION: PROTEIN CELLULAR FACTOR USEFUL FOR
/ NUMBER OF SEQUENCES: 7
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Arnold White & Durkee
/ STREET: P.O. Box 4433
/ CITY: Houston
/ STATE: Texas
/ COUNTRY: USA
/ ZIP: 77210
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patentin Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: PCT/US93/03077
/ FILING DATE: 19930331
/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US/07/862,025
/ FILING DATE: April 2, 1992
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Kammerer, Patricia A.
/ REGISTRATION NUMBER: 29,775
/ REFERENCE/DOCKET NUMBER: UTPD270PCT
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 713-787-1540
/ TELEFAX: 713-749-2679
/ TELEX:
/ INFORMATION FOR SEQ ID NO: 3:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 414 amino acids
/ TYPE: AMINO ACID
/ STRANDEDNESS: unknown
/ TOPOLOGY: unknown
/ MOLECULE TYPE: peptide
PCT-US93-03077-3

Query Match      15.0%; Score 106.5; DB 5; Length 414;
Best Local Similarity 22.4%; Pred. No. 0.029;
Matches 37; Conservative 33; Mismatches 50; Indels 45; Gaps 5;
QY 3 MANDAVLRLKRGKGAADQIIIEYLKQVSLKEKALIAQLTLEKKLRVENAKL 57

```

```

Db 1638 NKIYAEQKKEEKEEKKAEELKKEEKEEKKKA-EQLKKEEKEEKKVQQLKKEEKEEKK 1696
QY 64 QELIQAIONGVKQIAPPSGTPILANSVSENVIOSTAVTVSSGTGKQIGTGDEKKA 123
Db 1697 AEQJKEE-----ENKIK-----VEQLKKEEKEEKK 1724
QY 124 KEKIEKKEEKKKKQ 139
Db 1725 AEELKKEEKEEKKVQ 1740

RESULT 10
Q8IH3 PRELIMINARY; PRT; 2055 AA.
AC Q8IH3;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DE 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE MABBL, putative.
GN PF11_0486.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=22255705; PubMed=1236864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlson J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallow S.D., Suh B., Peterson J., Angiuoli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McPadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum."
RL Nature 419:498-511(2002).
DR EMBL; AE014843; AAN36066.1; --
SQ SEQUENCE 2055 AA; 243210 MW; 783F48BB7286B1F3 CRC64;

Query Match 15.7%; Score 111.5; DB 5; Length 2055;
Best Local Similarity 28.7%; Pred. No. 27;
Matches 39; Conservative 24; Mismatches 40; Indels 33; Gaps 3;

QY 4 NDAVLKRLKQKAGADQIIIEYLKQVSLKKEKAILQATLREKKLRVENAKLKKEIEELK 63
Db 1638 NKIYAEQKKEEKEEKKAEELKKEEKEEKKKA-EQLKKEEKEEKKVQQLKKEEKEEKK 1696
QY 64 QELIQAIONGVKQIAPPSGTPILANSVSENVIOSTAVTVSSGTGKQIGTGDEKKA 123
Db 1697 AEQJKEE-----ENKIK-----VEQLKKEEKEEKK 1724
QY 124 KEKIEKKEEKKKKQ 139
Db 1725 AEELKKEEKEEKKVQ 1740

RESULT 11
Q8IH3 PRELIMINARY; PRT; 964 AA.
AC Q8IH3;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DE 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Hypothetical protein.
GN PF11_0207.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=22255705; PubMed=1236864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlson J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallow S.D., Suh B., Peterson J., Angiuoli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McPadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum."
RL Nature 419:498-511(2002).
DR EMBL; AE014838; AAN35791.1; --
KW Hypothetical protein.
SQ SEQUENCE 964 AA; 116348 MW; 7EAC615834ED3B9 CRC64;

Query Match 15.7%; Score 111; DB 5; Length 964;
Best Local Similarity 25.0%; Pred. No. 13;
Matches 35; Conservative 43; Mismatches 42; Indels 20; Gaps 7;

QY 8 LKRLKQKAGADQIIIEYLKQVSLKKEKAILQATLRE-KKLRVENAKLKKEIEELKOE 65
Db 504 IKEIKKEIKKEIKKEIKKEIKKEIKKEIKKEIKKEIKKEIKKEIKKEIKKEIKKEIK 561
QY 66 L-IQAIONGVKQIAPPSGTPILANSVSENVIO-STAVTVSSGTGKQIGTGDEK 121
Db 562 IKEIKKEIKKEIKKEIKKEIKKEIKKEIKKEIKKEIKKEIKKEIKKEIKKEIKKEIK 611
QY 122 KAKEKI-EKKKEEKKKKQ 139
Db 612 EVKEIEKEIEKEIEKEIEKEIEKEIEKEIEKEIEKEIEKEIEKEIEKEIEKEIEKEIE 631

RESULT 12
Q8JH24 PRELIMINARY; PRT; 837 AA.
AC Q8JH24;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE GRIP-associated protein 1 long form.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley;
RX MEDLINE=2035053; PubMed=10896157;
RA Ye B., Liao D., Zhang X., Zhang P., Dong H., Haganir R.L.;
RA "GRASP-1: A Neuronal RasGEF Associated with the AMPA Receptor/GRIP
RT Complex."
RL Neuron 26:603-617(2000).
DR EMBL; AF274057; AAF82298.1; --
SQ SEQUENCE 837 AA; 96074 MW; A746AE4FD09D3AD2 CRC64;

Query Match 15.6%; Score 110.5; DB 11; Length 837;
Best Local Similarity 22.5%; Pred. No. 12;
Matches 47; Conservative 29; Mismatches 62; Indels 71; Gaps 6;

QY 1 MANDAVLKRLEQKAGADQIIIEYLKQVSLKKEKAILQATLRE-EKK 47
Db 31 LRKRGVLSLRQKVAVLDR-EFSKQKALSKSKQAQEVVLSSEKEMLOAKLHSQED 88
QY 48 LRVENAKLKKEIEELKQELIQAIONGVKQIAPPSGTPILANSVSENVIOSTAVTVSS 107
Db 89 FRLONSTLMAFSLKCSQLEQLLEENRQLKEGVGAAGPHVDBELRLQEAENTRALQKMA 148
QY 108 GTKEQ-----IKGTGD-----EKK 122

```

DB 149 ALGORYCAVRSVSEGGDPPGDLPLSLSPMLAEVLKMEEREKLLMEOLOG 208  
 QY 123 -----AKKIEKKKKKKQOS 140  
 DB 209 LESSKQATSRLOEELAKLSEKKKKQOS 237

RESULT 13  
 ID P90895 PRELIMINARY; PRT; 947 AA.  
 AC P90895; Q18791;  
 DT 01-MAY-1997 (TEMBLrel. 03, Created)  
 DT 01-MAY-1999 (TEMBLrel. 10, Last sequence update)  
 DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)  
 DE F57F5.2 protein.  
 GN F57F5.2.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peioderinae; Caenorhabditis.  
 NC NCB1\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Harris B., Holt R.;  
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; Z78012; CAB01418.1; -.  
 DR EMBL; Z75953; CAB01418.1; JOINED.  
 DR EMBL; Z75953; CAB00103.1; -.  
 DR EMBL; Z78012; CAB00103.1; JOINED.  
 DR HSSP; P26769; 1AB8.  
 DR WormPep; F57F5.2; CE20887.  
 DR InterPro; IPR001054; G\_cyclase.  
 DR Pfam; PF00211; guanylate\_cyc; 1.  
 DR SMART; SM00044; CYCC; 1.  
 DR PROSITE; PS00452; GUANYLATE\_CYCLASES\_1; 1.  
 DR PROSITE; PS50125; GUANYLATE\_CYCLASES\_2; 1.  
 DR Lyase.  
 KW K  
 SQ SEQUENCE 947 AA; 106998 MW; 947968427B146D64 CRC64;

Query Match 15.4%; Score 110; DB 5; Length 947;  
 Best Local Similarity 21.6%; Pred. No. 15;  
 Matches 37; Conservative 43; Mismatches 59; Indels 32; Gaps 6;

QY 1 MANNDVATKRLK---QKGAEDQIIEYLKQVSLKKEKALQATLREKKLRVENAKLKK 57  
 DB 741 LSSSDLAVSRSVSKSDGTPRTSELSKVRIRREALAQEK-BEERTTKENQKIEE 799  
 QY 58 EIEELKQE---LIQAEIQNGVKQIAF-----PSCGPLHANSVNSENVIQ 98  
 DB 800 VGEHDVSEATSLDSEVSHDNNISFQMPSPDSIPHEDRTSLSPATPSEIGDAISKKLE 859  
 QY 99 -----STAVTVSSGCKQIKGCTGDEKKAKIEKKKKKKQOSIAGS 145  
 DB 860 KEDNSNSWSSLDERTTVSAKPTT-----TRLLAQKLEKKKSSWASS 905

RESULT 14  
 ID Q9XUK4 PRELIMINARY; PRT; 244 AA.  
 AC Q9XUK4;  
 DT 01-NOV-1999 (TEMBLrel. 12, Created)  
 DT 01-NOV-1999 (TEMBLrel. 12, Last sequence update)  
 DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)  
 DE W07G1.1 protein.  
 GN W07G1.1.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peioderinae; Caenorhabditis.  
 NC NCB1\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Percy C.M.;  
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
 RN [2]

RP SEQUENCE FROM N.A.  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA none;  
 RT "Genome sequence of the nematode C.elegans: A platform for  
 RT investigating biology.";  
 RL Science 282:2012-2018 (1998).  
 DR EMBL; Z82076; CAB04933.1; -.  
 DR WormPep; W07G1.1; CE24035.  
 DR InterPro; IPR003533; DCX.  
 DR Pfam; PF03607; DCX; 1.  
 DR PROSITE; PS50309; DC; 1.  
 SQ SEQUENCE 244 AA; 27832 MW; 2F30B876F76FABIC CRC64;

Query Match 15.4%; Score 109.5; DB 5; Length 244;  
 Best Local Similarity 24.8%; Pred. No. 4;  
 Matches 37; Conservative 35; Mismatches 54; Indels 23; Gaps 5;

QY 10 RLEQKGAEDQIIE---YLKQVSL-LKEKALQATLREKKLRVENAKLKEIE 61  
 DB 75 KLKKKKKKKQTVQPTKEHIEIGDEISIPLEPVVLEKKKKKKKLQKNDLHGSIRT 134  
 QY 62 LKQEL---IQAEIQNGVKQIAFP---SCGPLHANSVNSENVIQSTAVTVSSGCKQIKG 115  
 DB 135 LKTAETELPSETBSEDEDEHHRRSATPDHHPATKHSVAPSRSSR-----RG 185  
 QY 116 GTGDEKKAKEKIEKKKKKKQOSIAGS 144  
 DB 186 SRRDEKREDEBERKKKDEKKKRSVSSS 214

RESULT 15  
 ID Q96216 PRELIMINARY; PRT; 2385 AA.  
 AC Q96216;  
 DT 01-MAY-1999 (TEMBLrel. 10, Created)  
 DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)  
 DE Hypothetical protein.  
 GN PFB0615C.  
 OS Plasmodium falciparum (isolate 3D7).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 NC NCB1\_TaxID=36329;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=99021743; PubMed=9804551;  
 RX Gardner M.J., Tetteelin H., Carucci D.J., Cummings L.M., Aravind L.,  
 RA Koonin E.V., Shalim S., Mason T., Yu K., Fujii C., Pederson J.,  
 RA Shen K., Jiang J., Ascon C., Lai Z., Schwartz D.C., Perera M.,  
 RA Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,  
 RA Fraser C.M., Adams M.D., Venter J.C., Hoffman S.L.;  
 RT "Chromosome 2 sequence of the human malaria parasite Plasmodium  
 RT falciparum.";  
 RL Science 282:1126-1132 (1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=3D7;  
 RX MEDLINE=2255705; PubMed=12368864;  
 RA Gardner M.J., Hall N., Fung B., White O., Berriman M., Hyman R.W.,  
 RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,  
 RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyse S.,  
 RA Chan M.-S., Nene V., Shalim S.J., Sub B., Peterson J., Angiolini S.,  
 RA Perera M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,  
 RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,  
 RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,  
 RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,  
 RA Fraser C.M., Barrall B.;  
 RT "Genome sequence of the human malaria parasite Plasmodium  
 RT falciparum.";  
 RL Nature 419:498-511 (2002).  
 DR EMBL; AE001406; AAC71912.2; -.  
 KW Hypothetical protein.  
 SQ SEQUENCE 2385 AA; 288299 MW; CC7CF642BCTEFOB9 CRC64;

Db 135 HNSNIKKLRKAKDKNENMVAKLKKVELELOHLQVLDGKEVEKQKHRENIKKLNS 194  
QY 58 EIEELKOEEL-----IOAEIONGVKQIAFPSTPLHANSWSENVIOSTA 101  
Db 195 MVEBOEKDGLQVDMDELEEKNSIOAALDSAYTEL-----TLHKHNAKDSBAQDA 249  
QY 102 VTVSSGTEQIKGDTGDEKAKEX---IEKGEKKEKKQOOSIA 142  
Db 250 LSR-----EMKAEELSALEKQAEBAQOQOETLA 279

## RESULT 9

PCT-US93-03077-1  
Sequence 1, Application PC/TUS9303077  
GENERAL INFORMATION:  
APPLICANT: Board of Regents, The University of Texas System  
APPLICANT: Gaynor, Richard B.  
APPLICANT: Wu, Foon Kin  
TITLE OF INVENTION: PROTEIN CELLULAR FACTOR USEFUL FOR  
TITLE OF INVENTION: REGULATING GENE EXPRESSION  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESS: Arnold White & Durkee  
STREET: P.O. Box 4433  
CITY: Houston  
STATE: Texas  
COUNTRY: USA  
ZIP: 77210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/03077  
FILING DATE: 19930331  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/862,025  
FILING DATE: April 2, 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Kammerer, Patricia A.  
REGISTRATION NUMBER: 29,775  
REFERENCE/DOCKET NUMBER: UTPD270PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 713-787-1540  
TELEFAX: 713-749-2679  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1093 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
PCT-US93-03077-1

Query Match 15.0%; Score 106.5; DB 5; Length 1093;  
Best Local Similarity 22.4%; Pred. No. 0.099;  
Matches 37; Conservative 33; Mismatches 50; Indels 45; Gaps 5;  
QY 3 NNDVAVLRLEKQKGAADQIEYLKQOVSILK-----KALIQATIREKKIARVERAKKK 57  
Db 571 HNSNIKKLRKAKDKNENMVAKLKKVELELOHLQVLDGKEVEKQKHRENIKKLNS 630  
QY 58 EIEELKOEEL-----IOAEIONGVKQIAFPSTPLHANSWSENVIOSTA 101  
Db 631 MVEBOEKDGLQVDMDELEEKNSIOAALDSAYTEL-----TLHKHNAKDSBAQDA 685  
QY 102 VTVSSGTEQIKGDTGDEKAKEX---IEKGEKKEKKQOOSIA 142  
Db 686 LSR-----EMKAEELSALEKQAEBAQOQOETLA 279

RESULT 10  
US-09-914-259-66  
Sequence 66, Application US/09914259  
Patent No. 6495336  
GENERAL INFORMATION:  
APPLICANT: Makowski, Lee  
APPLICANT: Hyman, Paul  
APPLICANT: Williams, Mark  
TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES  
FILE REFERENCE: 8471-010-999  
CURRENT APPLICATION NUMBER: US/09/914,259  
CURRENT FILING DATE: 2000-11-21  
NUMBER OF SEQ ID NOS: 180  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 66  
LENGTH: 961  
TYPE: PRT  
ORGANISM: Bos taurus  
US-09-914-259-66

Query Match 14.7%; Score 104; DB 4; Length 961;  
Best Local Similarity 23.3%; Pred. No. 0.14;  
Matches 42; Conservative 35; Mismatches 45; Indels 58; Gaps 7;

QY 1 MANDVAVLRLEKQKGAADQIEYLKQOVSILK-----EKALIQ-ATIREKK-----L 48  
Db 645 LEQHDSTVTHYKMNIRBDQLLEELKQOISTLKQNEQLOTAIVQVSOIQKHQDYNLL 704  
QY 49 RV-----ENAKLKEIEELK--QELIOAEIONGVKQIAFPS 82  
Db 705 KVGQKDSQHGCPYTDGQMNNGVQPEISRLREIEELKSRRLQQLAE----- 755  
QY 83 GTPLHANSWSENVIOSTAVTVSSGTEQIKGSTG- EKKAKEXIEKGEKKEKKQOOSI 141  
Db 756 -----KDSLIEMLKSSQSLPSTGNEQSSATATGDSQIAIEIKOELATILKSQLNQSV 805

RESULT 11  
US-08-903-801-1  
Sequence 1, Application US/08903801  
Patent No. 5932712  
GENERAL INFORMATION:  
APPLICANT: Hillman, Jennifer  
APPLICANT: Corley, Neil C.  
APPLICANT: Shah, Purvi  
TITLE OF INVENTION: NEW ANEXIN BINDING PROTEIN  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESS: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/903,801  
FILING DATE: Heretwlich  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0354 US  
TELECOMMUNICATION INFORMATION:

```

TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 290 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: PROSNONO1
CLONE: 2272281
US-08-903-801-1

Query Match 14.4% Score 102; DB 2; Length 290;
Best Local Similarity 27.0%; Pred. No. 0.049;
Matches 44; Conservative 18; Mismatches 57; Indels 44; Gaps 6

QY 13 OKGAADIIIVYKQV-----SLKKAIIQATL---REKKLR 49
Db 99 ESGESDSEFFLSRKQKKNQKNKPEPNIESGNEDDASFKITVAKKAERERKRD 158
QY 50 VENAKLK-KEIEEL---KQELIQAEIQNGVQIAPFSGTPIHANSWSENVISTAVT 103
Db 159 EKAKLRKTKKEKEHEFTGKQSQKXESQKPEE-----ETVKSATVDTGYI 206
QY 104 TVSSGTVEQIKGTGD--EKKAKKIKKGGKKKKQOOSNG 143
Db 207 PASEEKAETPTAAEDNDGDKKKDKKKKKGGKKEKEKEKKKG 249

RESULT 12
US-09-295-055-1
Sequence 1, Application US/09295055
Patent No. 6232440
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer
APPLICANT: Corley, Neil C.
APPLICANT: Shah, Putvi
TITLE OF INVENTION: NEW ANEXIN BINDING PROTEIN
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/295.055
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/903,801
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0354 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 290 amino acids
TYPE: amino acid
STRANDEDNESS: single

```

```
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: PROSCNON01  
; CLONE: 227281  
US-09-295-055-1
```

Query Match                14.4%; Score 102; DB 3; Length 290;  
Best Local Similarity 27.0%; Pred. No. 0.049;  
Matches 44; Conservative 16; Mismatches 57; Indels 44; Gaps 6;

Oy      13 QKGAEDQIIIEYLKQV-----SLLEKAILQATL---REEKL R 49  
        :|::||:|  
Db     99 ESGDESEDFIQRGQKNQNKGPNIESGNDDSDAFKITVAQKAERERRKRKD 158  
  
Oy      50 VENNKK--KEIEL----KOELIQAEIONGVQAIPPSGPLHANSVNVIQTAVT 103  
        |||||:||||  
Db     159 EEKKJAKRLKEKELETGGCKDOOSQOKESQRFEE-----ETVSKSVTVDTGYI 206  
  
Oy      104 TVSSGTKEQIKGTGD---EKKAKEKIEKGEKKEKKQSIAAG 143  
        |  
Db     207 PASEEKALPTPAEDDNEGDGKKKKKKKKKKKEKEKEKEKKKK 249

RESULT 13  
US-09-286-981B-3

Sequence 3, Application US/09286981B  
Patent No. 650351  
GENERAL INFORMATION:  
APPLICANT: Witzemann, Theresa M.  
APPLICANT: Koenig, Scott  
APPLICANT: Johnson, Leslie S  
TITLE OR INVENTION: Derivatives of Choline Binding Proteins for Vaccines  
FILE REFERENCE: 469201-396  
CURRENT APPLICATION NUMBER: US/09/286\_981B  
PRIOR FILING DATE: 1999-04-06  
PRIORITY APPLICATION NUMBER: US 60/085,743  
PRIOR FILING DATE: 1998-05-15  
NUMBER OF SEQ ID NOS: 38  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 3  
LENGTH: 431  
TYPE: PRF  
ORGANISM: Streptococcus pneumoniae  
US-09-286-981B-3

Query Match                14.4%; Score 102; DB 4; Length 431;  
Best Local Similarity 29.2%; Pred. No. 0.081;  
Matches 42; Conservative 25; Mismatches 57; Indels 20; Gaps 6;

Oy      1 MANDAVLVKRL-----EQKGAEDQIIIEYLKQQVSLLLEKAILQLTLREKKLRVENA 53  
        ::||:|  
Db     141 IASBDVVKKAAELLVXEAAGSHNEEKTKAKAAVESKKAFAATLERIKTERKKAEBEA 200  
  
Oy      54 KLKKEIFELMOELLQAEIOINGVQAIPPSGPLHANSNVSNVIQTAVTTVSSGTKEQI 113  
        :|::||:|  
Db     201 KRKAEEA---EEYNKLKRTKTGAF-GEP--ATPDKKENDAKSOSVKKSSKPIL 252  
  
Oy      114 KGCTDEKAKEKIEKGEKKEKK 137  
        |||||:||||  
Db     253 K-----SEKVAAE-AEKVAAEAKK 271

RESULT 14  
US-08-664-596B-11

Sequence 11, Application US/08664596B  
Patent No. 5607703  
GENERAL INFORMATION:  
APPLICANT: Jacobs, Kenneth  
APPLICANT: McCoy, John  
APPLICANT: Lavallee, Edward  
APPLICANT: Racie, Lisa  
APPLICANT: Merberg, David  
APPLICANT: Treacy, Maurice

APPLICANT: Evans, Cheryl  
 APPLICANT: Spaulding, Vikki  
 APPLICANT: Bowman, Michael  
 TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES  
 TITLE OF INVENTION: ENCODING THEM  
 NUMBER OF SEQUENCES: 37  
 CORRESPONDENCE ADDRESS:  
 ADDRESS: Genetics Institute, Inc.  
 STREET: 87 Cambridgepark Drive  
 CITY: Cambridge  
 STATE: Massachusetts  
 COUNTRY: U.S.A.  
 ZIP: 02140  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/664,596B  
 FILING DATE:  
 CLASSIFICATION: 514  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Brown, Scott A.  
 REGISTRATION NUMBER: 32,724  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617) 498-8224  
 TELEFAX: (617) 876-5851  
 INFORMATION FOR SEQ ID NO: 11:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 271 amino acids  
 TYPE: amino acid  
 STRANDEDNESS:  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-664-596B-11

Query Match 14.3%; Score 101.5; DB 1; Length 271;  
 Best Local Similarity 27.7%; Pred. No. 0.05;  
 Matches 39; Conservative 21; Mismatches 48; Indels 33; Gaps 7;  
 QY 1 MANDAVIKRLQKGAEDQIIEYIKQVSLKEKAILQATLREKKLRVENAKLKEIE 60  
 Db 131 MNSD-----LEKGA-----EINPKTALLPSDSVP-----AERNLKGLE 168  
 QY 61 ELKQELIOAEIQNGKQIAFPSCGTPHANSWSENVIOSTAVTYSSGTEQIKGTDE 120  
 Db 169 ESPSE--AEDPISGTTQTMVAEVAEVEKNETVSB-ILPSTCIVTLVPGIP-----TGDE 219  
 QY 121 KKAKEK--IEKKGEKKEKKQ 139  
 Db 220 KTVDKKNIKXKKGNMDEKKEK 240

RESULT 15  
 US-09-914-259-17  
 Sequence 17, Application US/09914259  
 Patent No. 6495336  
 GENERAL INFORMATION:  
 APPLICANT: Makowski, Lee  
 APPLICANT: Hyman, Paul  
 APPLICANT: Williams, Mark  
 TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES  
 FILE REFERENCE: 8471-010-999  
 CURRENT APPLICATION NUMBER: US/09/914,259  
 NUMBER OF SEQ ID NOS: 180  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 17  
 LENGTH: 956  
 TYPE: PRT  
 ORGANISM: Mus musculus  
 US-09-914-259-17

Query Match 14.2%; Score 101; DB 4; Length 956;  
 Best Local Similarity 25.0%; Pred. No. 0.27;  
 Matches 33; Conservative 30; Mismatches 55; Indels 14; Gaps 4;  
 QY 6 AVLRLEQKGAED--QIIEYIKQVSLKEKAILQATLREKK-----LIVENAKL 55  
 Db 422 SLVRLDQKDEINQSQSLAEKIKQ--MLDQDELLASTRDYKIQBELTRLQIENEA 479  
 QY 56 KKEIEELKQELIOAEIQNGKQIAFPSCGTPHANSWSENVIOSTAVTYSSGTEQIKG 115  
 Db 480 KDEVEVLAQLEELAVNYDQKQSEVEDKT--RANEQLTDELAQKTTTLTTQRELSQLQE 537  
 QY 116 GTGDEKKAKEXI 127  
 Db 538 LSNHQKKRATFI 549

Search completed: January 2, 2004, 18:23:40  
 Job time : 16.7176 secs

**THIS PAGE BLANK (USPTO)**



Run on: January 2, 2004, 18:22:28 ; Search time 32.8588 Seconds

(without alignments)  
897.605 Million cell updates/sec

Title: US-09-930-169-1  
 Perfect score: 709  
 Sequence: 1 MANNDAVLKRLIQKGAEDQ.....EKKGEEKKKOOSIAGSADS 147

Gapor 10.0 , Gapext 0.5

133937 seqs, 200641211 residues

total number of hits satisfying chosen parameters: 733937

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published Applications AA:\*

```

1:  Published Applications: Aa: *
2:  /cgn2_6/ptodata/1/pubppaa/US07_PUBCOMB.pcp.*
3:  /cgn2_6/ptodata/1/pubppaa/PCT_NEW_PUB.pcp.*
4:  /cgn2_6/ptodata/1/pubppaa/US06_NEW_PUB.pcp.*
5:  /cgn2_6/ptodata/1/pubppaa/US07_PUBCOMB.pcp.*
6:  /cgn2_6/ptodata/1/pubppaa/US07_NEW_PUB.pcp.*
7:  /cgn2_6/ptodata/1/pubppaa/PCTUS_PUBCOMB.pcp.*
8:  /cgn2_6/ptodata/1/pubppaa/US08_NEW_PUB.pcp.*
9:  /cgn2_6/ptodata/1/pubppaa/US08_PUBCOMB.pcp.*
10: /cgn2_6/ptodata/1/pubppaa/US09_PUBCOMB.pcp.*
11: /cgn2_6/ptodata/1/pubppaa/US09C_PUBCOMB.pcp.*
12: /cgn2_6/ptodata/1/pubppaa/US09_NEW_PUB.pcp.*
13: /cgn2_6/ptodata/1/pubppaa/US10_PUBCOMB.pcp.*
14: /cgn2_6/ptodata/1/pubppaa/US10B_PUBCOMB.pcp.*
15: /cgn2_6/ptodata/1/pubppaa/US10C_PUBCOMB.pcp.*
16: /cgn2_6/ptodata/1/pubppaa/US10_NEW_PUB.pcp.*
17: /cgn2_6/ptodata/1/pubppaa/US60_NEW_PUB.pcp.*
18: /cgn2_6/ptodata/1/pubppaa/US60_PUBCOMB.pcp.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result	Score	Query	Match	Length	DB	ID	Description
No.							
1	709	100.0	147	11	US-09-930-169-1		Sequence 1, Appl1
2	704	99.3	312	10	US-09-883-026-36		Sequence 36, Appl
3	516	72.8	108	11	US-09-930-169-2		Sequence 2, Appl
4	509	71.8	310	10	US-09-883-026-35		Sequence 35, Appl
5	275	38.9	168	11	US-09-930-169-3		Sequence 3, Appl
6	147	20.7	56	10	US-09-813-718-22		Sequence 22, Appl
7	116.5	15.4	1079	12	US-10-369-493-116		Sequence 216, Appl
8	107	16.1	1365	12	US-10-369-493-3279		Sequence 3279, Appl
9	106.5	15.0	1069	15	US-10-146-473-777		Sequence 77, Appl
10	105.5	14.9	1884	10	US-09-785-7704-17		Sequence 17, Appl
11	105.5	14.9	1307	10	US-09-785-7704-16		Sequence 16, Appl
12	105	14.8	670	11	US-09-298-5538-63		Sequence 63, Appl
13	104.5	14.7	1231	12	US-10-369-493-3503		Sequence 503, Appl
14	104	14.7	711	11	US-09-298-5538-3		Sequence 3, Appl
15	104	14.7	961	12	US-10-080-608A-66		Sequence 66, Appl

16	104	14.7	961	12	US-10-370-685-155	Sequence 155, App
17	103.5	14.6	496	11	US-09-298-523B-4	Sequence 4, Appl
18	103	14.5	286	10	US-09-881-752B-352	Sequence 352, A
19	103	14.5	1130	12	US-10-369-493-6751	Sequence 6751, A
20	103	14.5	1169	12	US-10-369-493-6751	Sequence 1095, A
21	102	14.4	290	10	US-09-808-883-1	Sequence 1, Appl
22	102	14.4	431	12	US-10-354-995-3	Sequence 3, Appl
23	101.5	14.3	313	14	US-10-114-893-90	Sequence 90, App
24	101.5	14.3	890	12	US-09-849-602-19	Sequence 17, App
25	101	14.2	956	12	US-10-080-608A-17	Sequence 11, App
26	101	14.2	956	12	US-10-370-685-106	Sequence 106, App
27	101	14.2	957	12	US-10-080-608A-16	Sequence 16, App
28	101	14.2	957	12	US-10-370-685-105	Sequence 105, App
29	100.5	14.2	419	12	US-10-370-685-105	Sequence 15, App
30	100.5	14.2	433	11	US-09-056-019-33	Sequence 33, App
31	100.5	14.2	690	11	US-09-298-523B-61	Sequence 61, App
32	100.5	14.2	701	11	US-09-298-523B-62	Sequence 62, App
33	100	14.1	589	11	US-09-298-523B-14	Sequence 14, App
34	100	14.1	691	11	US-09-298-523B-1	Sequence 1, App
35	100	14.1	707	11	US-09-298-523B-2	Sequence 2, App
36	100	14.1	929	11	US-09-298-523B-2	Sequence 60, App
37	99.5	14.0	886	12	US-10-369-493-1016	Sequence 1016, App
38	99	14.0	2139	9	US-09-727-384-6	Sequence 6, Appl
39	99	14.0	2139	9	US-10-022-219-4	Sequence 4, Appl
40	98.5	13.9	1179	12	US-10-369-493-3330	Sequence 1330, A
41	98.5	13.9	1179	12	US-10-369-493-20394	Sequence 20394, A
42	98.5	13.9	1837	12	US-10-369-493-5368	Sequence 5368, App
43	98	13.8	374	9	US-09-925-302-711	Sequence 711, App
44	98	13.8	1203	15	US-10-097-340-43	Sequence 43, App
45	97.5	13.8	1164	12	US-10-369-493-9770	Sequence 9770, App

## ALIGNMENTS

his case

RESULT 1  
 US-09-930-169-1  
 : Sequence 1, Application US/099330169  
 : Publication No. US20030004309a1  
 : GENERAL INFORMATION:  
 : APPLICANT: KIM, SUNGHOON  
 : APPLICANT: KO, YOUNG-GYU  
 : TITLE OF INVENTION: IMMUNOLOGICAL ENHANCEMENT AGENT COMPRISING N-TERMINAL PEPTIDE OF  
 : FILE REFERENCE: 058335/0106 P43 AS AN EFFECTIVE COMPONENT  
 : CURRENT APPLICATION NUMBER: US/09/930, 169 *this case*  
 : PRIOR FILING DATE: 2001-08-16  
 : PRIOR APPLICATION NUMBER: KR 2001-91310  
 : NUMBER OF SEQ ID NOS: 11  
 : SOFTWARE: patentin Ver. 2.1  
 : SEQ ID NO 1  
 : LENGTH: 147  
 : TYPE: PRT  
 : ORGANISM: Unknown Organism  
 : FEATURE:  
 : OTHER INFORMATION: Description of Unknown Organism: Mammalian  
 : OTHER INFORMATION: protein sequence  
 : US-09-930-169-1

Query Match	100.0%	Score 709;	DB 11;	Length 147;
Best Local Similarity	100.0%;	Prod. No. 7.7e-54;		
Matches 147;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MANNDAVLRLEFOKGAERDQIIIEYLKQOVSLLERKAILDATTREKKRLRVENAKLKEIE	60	
Db	1	MANNDAVLRLEFOKGAERDQIIIEYLKQOVSLLERKAILDATTREKKRLRVENAKLKEIE	60	
QY	61	ELKQELIDAEIONGVKQIAPPSGPHIAMSMSENVIOSTATTYSSGTEQIKGGTDE	120	
Db	61	ELKQELIDAEIONGVKQIAPPSGPHIAMSMSENVIOSTATTYSSGTEQIKGGTDE	120	
QY	121	KKAEKIEIKKEKKERKQOOSIAGSADS	147	

Db 121 KKAKETKEKKEKKOQSIAGSADS 147

RESULT 2  
US-09-851-026-36

Sequence 36, Application US/09851026  
Patent No. US20020160957A1

GENERAL INFORMATION:

APPLICANT: Stern, David M.  
Klaus, Matthias

Kao, Janet

Libutti, Steven K

TITLE OF INVENTION: Endothelial Monocyte Activating  
Polypeptide II: A Mediator Which Activates Host Response

NUMBER OF SEQUENCES: 42

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham, LLP

STREET: 1185 Avenue of the Americas  
CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10036

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.30, Version #1.30

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/851,026

FILING DATE: 07-May-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/360,821

FILING DATE: 08-OCT-96

ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.

REGISTRATION NUMBER: 28,678

REFERENCE/DOCKET NUMBER: 41735-A-PCT-US

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-278-0400

TELEFAX: 212-391-0525

INFORMATION FOR SEQ ID NO: 36:

SEQUENCE CHARACTERISTICS:  
LENGTH: 312 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 36:

US-09-851-026-36

Query Match 99.3%; Score 704; DB 10; Length 312;

Best Local Similarity 99.3%; Pred. No. 5.3e-53;

Matches 146; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 MANNDAVILKRLLEQKGAADQIIIEYLKQOVSLKEKAILOATLREKKLRVENAKLKEIE 60

1 MANNDAVILKRLLEQKGAADQIIIEYLKQOVSLKEKAILOATLREKKLRVENAKLKEIE 60

61 ELKQELIQAEIQNGVKQIAPPSGTPPLHANSVSENVIOSTAVTVSSGTRKQIKSGTGE 120

61 ELKQELIQAEIQNGVKQIAPPSGTPPLHANSVSENVIOSTAVTVSSGTRKQIKSGTGE 120

121 KKAKETKEKKEKKOQSIAGSADS 147

121 KKAKETKEKKEKKOQSIAGSADS 147

Publication No. US20030004309A1  
GENERAL INFORMATION:  
APPLICANT: KIM, SUNGHOON  
APPLICANT: KO, YOUNG-GYU  
TITLE OF INVENTION: IMMUNOLOGICAL ENHANCEMENT AGENT COMPRISING N-TERMINAL PEPTIDE OF  
TITLE OF INVENTION: P43 AS AN EFFECTIVE COMPONENT  
FILE REFERENCE: 058333/0106  
CURRENT APPLICATION NUMBER: US/09/930,169  
CURRENT FILING DATE: 2001-08-16  
PRIOR APPLICATION NUMBER: KR 2001-31310  
PRIOR FILING DATE: 2001-06-05  
NUMBER OF SEQ ID NOS: 11  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 2  
LENGTH: 108  
TYPE: PRT  
ORGANISM: Unknown Organism  
OTHER INFORMATION: Description of Unknown Organism: Mammalian  
OTHER INFORMATION: protein sequence

US-09-930-169-2

Query Match 72.8%; Score 516; DB 11; Length 108;

Best Local Similarity 100.0%; Pred. No. 2.4e-37;

Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MANNDAVILKRLLEQKGAADQIIIEYLKQOVSLKEKAILOATLREKKLRVENAKLKEIE 60

1 MANNDAVILKRLLEQKGAADQIIIEYLKQOVSLKEKAILOATLREKKLRVENAKLKEIE 60

61 ELKQELIQAEIQNGVKQIAPPSGTPPLHANSVSENVIOSTAVTVSSG 108

61 ELKQELIQAEIQNGVKQIAPPSGTPPLHANSVSENVIOSTAVTVSSG 108

RESULT 4

US-09-851-026-35

Sequence 35, Application US/09851026

Patent No. US20020160957A1

GENERAL INFORMATION:

APPLICANT: Stern, David M.  
Klaus, Matthias

Kao, Janet

Libutti, Steven K

TITLE OF INVENTION: Endothelial Monocyte Activating  
Polypeptide II: A Mediator Which Activates Host Response

NUMBER OF SEQUENCES: 42

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham, LLP

STREET: 1185 Avenue of the Americas  
CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10036

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.30, Version #1.30

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/851,026

FILING DATE: 07-May-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/360,821

FILING DATE: 08-OCT-96

ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.

REGISTRATION NUMBER: 28,678

REFERENCE/DOCKET NUMBER: 41735-A-PCT-US

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-278-0400

US-09-851-026-35

Query Match 72.8%; Score 516; DB 11; Length 108;

Best Local Similarity 100.0%; Pred. No. 2.4e-37;

Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MANNDAVILKRLLEQKGAADQIIIEYLKQOVSLKEKAILOATLREKKLRVENAKLKEIE 60

1 MANNDAVILKRLLEQKGAADQIIIEYLKQOVSLKEKAILOATLREKKLRVENAKLKEIE 60

61 ELKQELIQAEIQNGVKQIAPPSGTPPLHANSVSENVIOSTAVTVSSG 108

TELERAX: 212-391-0525  
INFORMATION FOR SEQ ID NO: 35:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 310 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 35:  
US-09-851-026-35

Query Match 71.8%; Score 509; DB 10; Length 310;  
Best Local Similarity 76.4%; Pred. No. 3,66-36;  
Matches 113; Conservative 8; Mismatches 23; Indels 4; Gaps 3;

Qy 1 MANDAVAKRLKQKGAADQIIEYKQVSLKKAIIQATLREKTLRVENAKLKEIE 60  
Db 1 MATNDVAVKRLKQKGAADQIIEYKQVSLKKAIIQATLREKTLRVENAKLKEIE 60  
Qy 61 ELKQELIAEIQNGVQKQIAPPSGTEPLHANSVSENVIOSTAV-TTVSSGTEKQIKGTGD 119  
Db 61 ELKQELIAEIQNGVQKQIAPPSGTEPLHANSVSENVIOSTAV-TTVSSGTEKQIKGTGD 118  
Qy 120 EKKAKETKEKKEKKEKQKQSIAGSADS 147  
Db 119 EKKAKETKEKKEKKEKQKQSIAGSADS 145

RESULT 5  
US-09-930-169-3  
Sequence 3, Application US/09930169  
Publication No. US200300043092A1  
GENERAL INFORMATION:  
APPLICANT: KIM, SUNGHOO  
APPLICANT: KO, YOUNG-GYU  
TITLE OF INVENTION: IMMUNOLOGICAL ENHANCEMENT AGENT COMPRISING N-TERMINAL PEPTIDE OF  
TITLE OF INVENTION: P43 AS AN EFFECTIVE COMPONENT  
FILE REFERENCE: 058333/0106  
CURRENT APPLICATION NUMBER: US/09/930,169  
CURRENT FILING DATE: 2001-08-16  
PRIOR APPLICATION NUMBER: KR 2001-31310  
PRIOR FILING DATE: 2001-06-05  
NUMBER OF SEQ ID NOS: 11  
SOFTWARE: Patent In Ver. 2.1  
SEQ ID NO 3  
LENGTH: 166  
TYPE: PRT  
ORGANISM: Unknown Organism  
FEATURE:  
OTHER INFORMATION: Description of Unknown Organism: Mammalian  
US-09-930-169-3

Query Match 38.8%; Score 275; DB 11; Length 166;  
Best Local Similarity 100.0%; Pred. No. 2,7e-16;  
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 91 MVSENVIOSTAVTVSSGTEKQIKGTGDEKKAKETKEKKEKQKQSIAGSADS 147  
Db 1 MVSENVIOSTAVTVSSGTEKQIKGTGDEKKAKETKEKKEKQKQSIAGSADS 57

RESULT 6  
US-09-813-718-22  
Sequence 22, Application US/09813718  
Publication No. US2002018266A1  
GENERAL INFORMATION:  
APPLICANT: Schimmel, Paul  
APPLICANT: Makasui, Keisuke  
TITLE OF INVENTION: Human Aminoacyl-tRNA Synthetase Polypeptides Useful For  
TITLE OF INVENTION: The Regulation of Angiogenesis  
FILE REFERENCE: 00-221  
CURRENT APPLICATION NUMBER: US/09/813,718

CURRENT FILING DATE: 2001-03-21  
NUMBER OF SEQ ID NOS: 58  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 22  
LENGTH: 58  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-813-718-22

Query Match 20.7%; Score 147; DB 10; Length 58;  
Best Local Similarity 100.0%; Pred. No. 8,2e-06;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 118 GDEKAKETKEKKEKKEKQKQSIAGSADS 147  
Db 1 GDEKAKETKEKKEKKEKQKQSIAGSADS 30

RESULT 7  
US-10-369-493-2116  
Sequence 2116, Application US/10369493  
Publication No. US20030233675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xianfeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 2116  
LENGTH: 1079  
TYPE: PRT  
ORGANISM: Schizosaccharomyces pombe  
US-10-369-493-2116

Query Match 16.4%; Score 116.5; DB 12; Length 1079;  
Best Local Similarity 21.6%; Pred. No. 0.13;  
Matches 38; Conservative 41; Mismatches 58; Indels 39; Gaps 5;

Qy 3 NNDVAVKRLKQKGAADQIIEYKQVSLKKAII-----QATLREKTLRVENA 53  
Db 246 NNTALQKMLEKRAEERIRREERARIAEERKRLAEVERARKEARLKKERKKKE 305  
Qy 54 KLR-----KEIHELQELQAEIQNGVQKQIAPPSG---TPLHN----- 89  
Db 306 EMKAQKYLKSKQKEQQLAQRRLQOMLESGVRVAGLSNEKKQKPYTNKKKNREGTS 365  
Qy 90 MVSENVIOSTAVTVSSGTEKQIKGTGDEKKAKETKEKKEKQKQSIAGSADS 145  
Db 366 SISSGILSESPNISIVDERQK-----DSKDSKVEKETEVEKRENEAEAA 415

RESULT 8  
US-10-369-493-3279  
Sequence 3279, Application US/10369493  
Publication No. US20030233675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xianfeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493

;; CURRENT FILING DATE: 2003-02-28  
;; PRIOR APPLICATION NUMBER: US 60/360,039  
;; PRIOR FILING DATE: 2002-02-21  
;; NUMBER OF SEQ ID NOS: 47374  
;; SEQ ID NO 3279  
;; LENGTH: 1965  
;; TYPE: PRT  
;; ORGANISM: Neurospora crassa  
US-10-369-493-3279

Query Match 15.1%; Score 107; DB 12; Length 1965;  
Best Local Similarity 25.0%; Pred. No. 1.8;  
Matches 44; Conservative 35; Mismatches 53; Indels 44; Gaps 9;

QY 2 ANNDVLRLEKQKGAEDQIIEYLKQOVSLLKEKALIQATLREKKLRVENAKLKKKEIE-60  
DB 1502 AEPEAKADLEER--ADALVKORSEKMTALNKLKLAESKEAMKOTQDERQKLONEYDI-1558  
QY 61 ELKQEL-IOAEIQNGVKQIAPF-----SGTF-----LHANSWV-92  
DB 1559 KIQQELAILKAEQSGVSGVGPATPAKPPTAQTQPGAGTGLPDLNLTDQTRRELI-1618  
QY 93 SENVIQSTAVTVSSGTEQIKGGTDEKAKKIE--KKGEKKK--KQOSIA-142  
DB 1619 SSNOV---IMGIKSNVKKNI---AAESKKVRELEVVVAEYOKIASAKEQAAA-1668

RESULT 9  
US-10-146-473-77  
Sequence 77, Application US/10146473  
Publication No. US20030108888A1

;; GENERAL INFORMATION:  
;; APPLICANT: Scanlan, Matthew  
;; APPLICANT: Scout, Ivan  
;; APPLICANT: Stockert, Elisabeth  
;; APPLICANT: Gure, Ali  
;; APPLICANT: Chen, Yao-Tseeng  
;; APPLICANT: Old, Lloyd  
;; TITLE OF INVENTION: Breast Cancer Antigens  
;; FILE REFERENCE: L00461/70130(JRY)  
;; CURRENT APPLICATION NUMBER: US/10/146,473  
;; CURRENT FILING DATE: 2002-05-15  
;; PRIOR APPLICATION NUMBER: US 60/291,150  
;; PRIOR FILING DATE: 2001-05-15  
;; NUMBER OF SEQ ID NOS: 82  
;; SOFTWARE: Patentin version 3.0  
;; SEQ ID NO 77  
;; LENGTH: 1069  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; NAME/KEY: UNSURE  
;; LOCATION: (268)..  
;; OTHER INFORMATION: X = any amino acid  
US-10-146-473-77

Query Match 15.0%; Score 106.5; DB 15; Length 1069;  
Best Local Similarity 22.4%; Pred. No. 0.96;  
Matches 37; Conservative 33; Mismatches 50; Indels 45; Gaps 5;

QY 3 NNDVLRLEKQKGAEDQIIEYLKQOVSLLKE-----KALIQATLREKKLRVENAKLKK-57  
DB 564 HNGNIITIKLPAKKNENMVAKANKYKELFEELQHILKQVLDGKEVEKQHRNIIKILNS-623  
QY 58 EIEELKQEL-IOAEIQNGVKQIAPFSGTPLHANSWVSENVIQSTIA-101  
DB 624 MVRQEKDLRLQVMDLEEKRSIOALDSAYKEL-----TDLHANAAXOSEQAAA-678  
QY 102 VTIIVSSGTKEQIKGGTDEKAKKIE--LEKGEKKKQOSIA-142  
DB 679 LSR-----EMKAKEELSALEKQAEKARQOQETIA-708

RESULT 10  
US-09-785-770A-17  
Sequence 17, Application US/09785770A  
Patent No. US20020103360A1

;; GENERAL INFORMATION:  
;; APPLICANT: Pan, Yang  
;; APPLICANT: Barnes, Thomas M.  
;; TITLE OF INVENTION: A NOVEL PROTEIN RELATED TO MELANOMA-INHIBITING PROTEIN  
;; FILE REFERENCE: 07334-328001  
;; CURRENT APPLICATION NUMBER: US/09/785,770A  
;; CURRENT FILING DATE: 2001-02-16  
;; PRIOR APPLICATION NUMBER: US 09/387,462  
;; PRIOR FILING DATE: 1999-09-01  
;; PRIOR APPLICATION NUMBER: US 09/145,056  
;; PRIOR FILING DATE: 1998-09-01  
;; NUMBER OF SEQ ID NOS: 24  
;; SOFTWARE: Patentin Ver. 4.0  
;; SEQ ID NO 17  
;; LENGTH: 1884  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-09-785-770A-17

Query Match 14.9%; Score 105.5; DB 10; Length 1884;  
Best Local Similarity 26.1%; Pred. No. 2.4;  
Matches 41; Conservative 30; Mismatches 55; Indels 31; Gaps 6;

QY 16 APADQIIEYLKQOVSLLKEK-----ALIQATLREK-----KLRVENAKLKEIEELK-63  
DB 1282 SENKSIIEKLKQVISMNASEFSEVOIALNEAKLSEKVSCHGRVQEEENARLKKKEQLO-1341  
QY 64 QEL-----IOAEIQNGVK-----QIAFSGTPLHANSWVSENVIQSTAVTVSSGT-109  
DB 1342 QEIEDMSKHAELSEQIISFEFSQKLEVAL--THKDDNIALTNCTIQLNLBCESES-1398  
QY 110 KEQIKGTGDEKAKKIEKKGEKKKQOSIAGSAD-146  
DB 1399 EGQNGGNDSDLANGEV--GGDRNEKKNQIKQMMND-1433

RESULT 11  
US-09-785-770A-16  
Sequence 16, Application US/09785770A  
Patent No. US20020103360A1

;; GENERAL INFORMATION:  
;; APPLICANT: Pan, Yang  
;; APPLICANT: Barnes, Thomas M.  
;; TITLE OF INVENTION: A NOVEL PROTEIN RELATED TO MELANOMA-INHIBITING PROTEIN  
;; FILE REFERENCE: 07334-328001  
;; CURRENT APPLICATION NUMBER: US/09/785,770A  
;; CURRENT FILING DATE: 2001-02-16  
;; PRIOR APPLICATION NUMBER: US 09/387,462  
;; PRIOR FILING DATE: 1999-09-01  
;; PRIOR APPLICATION NUMBER: US 09/145,056  
;; PRIOR FILING DATE: 1998-09-01  
;; NUMBER OF SEQ ID NOS: 24  
;; SOFTWARE: Patentin Ver. 4.0  
;; SEQ ID NO 16  
;; LENGTH: 1907  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-09-785-770A-16

Query Match 14.9%; Score 105.5; DB 10; Length 1907;  
Best Local Similarity 26.1%; Pred. No. 2.4;  
Matches 41; Conservative 30; Mismatches 55; Indels 31; Gaps 6;

QY 16 APADQIIEYLKQOVSLLKEK-----ALIQATLREK-----KLRVENAKLKEIEELK-63  
DB 1305 SENKSIIEKLKQVISMNASEFSEVOIALNEAKLSEKVSCHGRVQEEENARLKKKEQLO-1364

QY 64 QEL-----IQAEIQNGVK-----QIAPSGTPLHANSVSNVIOSTAVTTSSGT 109  
Db 1365 QEIDWSTLHAEISFOIKSEFKSQDLEVAL---THKDNINATNCITQNLLECESES 1421  
QY 110 KEQIKGTGDEKAKKEIKEXKEKKQKQSIAGSAD 146  
Db 1422 EGQNGKGNDSBLANGEV--GGDRNEKMKNOIKQMD 1456

## RESULT 12

US-09-298-523B-63  
; Sequence 63, Application US/09298523B  
; Publication No. US20030059438A1  
; GENERAL INFORMATION:  
; APPLICANT: BRILES et al.  
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS  
; FILE REFERENCE: 454312-3140  
; CURRENT FILING DATE: 1999-04-23  
; NUMBER OF SEQ ID NOS: 78  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 63  
; LENGTH: 670  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-298-523B-63

Query Match 14.8%; Score 105; DB 11; Length 670;  
Best Local Similarity 27.1%; Pred. No. 0.72;  
Matches 38; Conservative 27; Mismatches 51; Indels 24; Gaps 5;

QY 12 EOKGAADQIIEYKQVSLKEKAILQATLREEKLRVE---NAKL-KKEIEELKQEL 67  
Db 356 EKVVAEAKKVEAKKAKKQKQKEDRRNYPTNTYKTLLEAESDVKKVKAELIVKEEAA 415  
QY 68 QAEIQNGVKQIAPSGTPLHANSVSNVIOSTAVTTSSGTKEQIKGTGDEKAKKEKI 127  
Db 416 ESRNEEKIKQ-----AKEKVESKKAELATRL-----EKIK--TDRKKAEEBA 455  
QY 128 EKKEKKKKKQKQSIAGSADS 147  
Db 456 KRKAEESEKKAELAKQKQVDA 475

## RESULT 13

US-10-369-493-3503  
; Sequence 3503, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; PRIOR FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 3503  
; LENGTH: 1231  
; TYPE: PRT  
; ORGANISM: Neurospora crassa

FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(1231)  
; OTHER INFORMATION: unsure at all Xaa locations  
US-10-369-493-3503

Query Match 14.7%; Score 104.5; DB 12; Length 1231;  
Best Local Similarity 23.6%; Pred. No. 1.7;  
Matches 35; Conservative 28; Mismatches 56; Indels 29; Gaps 3;

QY 23 EYLQVQVSLKEKAILQATLREEKLRVENAKLKEIEELKQEL--IQA-----EIQNGVK 76  
Db 252 DYVQOEKTLQOSADLEKQKQRELESARLKNESHSLEEDLQRYVAQORDKEIKKQK 311  
QY 77 QIAPSGTPLHANSV-----SENVIOSTAVTTSSGTKEQI 113  
Db 312 AQALBEAVKQIANELVRLATVVDLKKSSMKEQERRRAGEQNVADLEALAKETTYEKI 371  
QY 114 KGTGDEKAKKEIKEXKEKKQKQSI 141  
Db 372 KAKYDAKETLEKQREQEDTKEELLQTL 399

## RESULT 14

US-09-298-523B-3  
; Sequence 3, Application US/09298523B  
; Publication No. US20030059438A1  
; GENERAL INFORMATION:  
; APPLICANT: BRILES et al.  
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS  
; FILE REFERENCE: 454312-3140  
; CURRENT APPLICATION NUMBER: US/09/298,523B  
; CURRENT FILING DATE: 1999-04-23  
; NUMBER OF SEQ ID NOS: 78  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 711  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-298-523B-3

Query Match 14.7%; Score 104; DB 11; Length 711;  
Best Local Similarity 27.0%; Pred. No. 0.95;  
Matches 38; Conservative 27; Mismatches 50; Indels 26; Gaps 5;

QY 12 EOKGAADQIIEYKQVSLKEKAILQATLREEKLRVE---NAKLK-EIEELKQEL 66  
Db 397 EKVVAEAKKVEAKKAKKQKQKEDRRNYPTNTYKTLLEAESDVKKVKAELIVKEEBA 456  
QY 67 IQAEIQNGVKQIAPSGTPLHANSVSNVIOSTAVTTSSGTKEQIKGTGDEKAKKEK 126  
Db 457 NESRNEEKIKQ-----AKEKVESKKAELATRL-----EKIK--TDRKKAEEBA 495  
QY 127 IEKKEKKKKKQKQSIAGSADS 147  
Db 496 AKRKAEESEKKAELAKQKQVDA 516

## RESULT 15

US-10-080-608A-66  
; Sequence 66, Application US/10080608A  
; Publication No. US20030198956A1  
; GENERAL INFORMATION:  
; APPLICANT: Makowski, Lee  
; APPLICANT: Hyman, Paul  
; APPLICANT: Williams, Mark  
; TITLE OF INVENTION: STRAGED ASSEMBLY OF NANOSTRUCTURES  
; FILE REFERENCE: 8471-010-999  
; CURRENT APPLICATION NUMBER: US/10/080,608A  
; CURRENT FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 180  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 66  
; LENGTH: 961  
; TYPE: PRT  
; ORGANISM: Bos taurus

US-10-080-608A-66



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2004, 18:16:37 ; Search time 32.1882 Seconds  
(without alignments)  
532.569 Million cell updates/sec

Title: US-09-930-169-2

Perfect score: 516  
Sequence: 1 MANNDAVLKRLKQKAEADQ.....NSWSENVISTAVTVSSG 108

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*  
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	516	100.0	312	23	ABB07442
2	516	100.0	328	22	AAB63244
3	395	76.6	310	16	AAB72577
4	106	20.5	294	22	ABBS9210
5	100.5	19.5	672	21	AAB21233
6	96	18.6	1975	22	ABBE2094
7	93	18.0	534	19	AAW46823
8	93	18.0	534	20	AAW89948
9	92.5	17.9	1047	22	AAW78507

10	92.5	17.9	1047	22	AAW79491
11	90	17.4	471	21	AAB41643
12	89	17.2	856	22	AAU35717
13	89	17.2	1177	22	AAAB6721
14	88.5	17.2	472	22	ABG06854
15	88	17.1	667	21	AAOI7717
16	87.5	17.0	345	23	ABBS7304
17	87.5	17.0	459	21	AAOI7716
18	87.5	17.0	477	21	AAOI7712
19	87	16.9	346	19	AAW60569
20	87	16.9	346	24	ABP72351
21	87	16.9	1305	22	ABBS8287
22	86.5	16.8	574	22	AABS5497
23	86.5	16.8	808	22	ABG05140
24	86.5	16.8	931	22	AAW79504
25	86.5	16.8	957	23	ABP69429
26	86.5	16.8	990	22	AAW78500
27	86	16.7	1024	23	ABBS7200
28	85.5	16.6	484	22	AAW78985
29	85.5	16.6	533	22	AAW79969
30	85.5	16.6	972	24	ABJ19394
31	85	16.5	223	21	AABS7038
32	85	16.5	1066	22	ABBS4251
33	84.5	16.4	377	23	ABBS0576
34	84	16.3	135	22	AAO03628
35	84	16.3	336	23	ABBS7804
36	84	16.3	502	22	AAW38736
37	84	16.3	591	22	AAW38737
38	84	16.3	2135	23	AAE21714
39	84	16.3	2382	24	ABP71619
40	83.5	16.2	242	21	AAAG10303
41	83.5	16.2	365	21	AAAG10302
42	83.5	16.2	383	21	AAOI0301
43	83.5	16.2	1122	21	ABU70418
44	83.5	16.2	1193	22	AAW5602
45	83.5	16.2	1193	23	ABG61824

#### ALIGNMENTS

RESULT 1	ID	ABB07442	standard; Protein; 312 AA.
XX	XX	ABB07442:	
AC	AC	ABB07442:	
XX	XX	23-APR-2002 (first entry)	
DT	DT	23-APR-2002 (first entry)	
XX	XX	Human p43 polypeptide.	
DE	DE	Human p43 polypeptide.	
XX	XX	Human; p43; cytosolic; anti-tumour; anti-angiogenic; EMAP II;	
KW	KW	cytokine; chemotaxis.	
XX	XX		
OS	OS	Homo sapiens.	
XX	XX		
FT	FT	Key	Location/Qualifiers
FT	FT	Domain	1..146
FT	FT	Domain	/note= "N-terminal domain"
FT	FT	Domain	147..312
FT	FT	Domain	/note= "C-terminal domain"
FT	FT	Region	10..21
FT	FT	Region	/note= "beta1 strand"
FT	FT	Region	28..34
FT	FT	Region	/note= "beta2 strand"
FT	FT	Region	40..46
FT	FT	Region	/note= "beta3 strand"
FT	FT	Region	53..56
FT	FT	Region	/note= "alpha 1 helix"
FT	FT	Region	59..66
FT	FT	Region	/note= "beta4 strand"
FT	FT	Region	70..72
FT	FT	Region	/note= "beta5 strand"

Human protein SEQ  
Human ORF ORF1407  
Helicobacter pylori  
Putative P. abyssi  
Novel human diagno  
O cuticulus endocri  
Novel human protea  
Human protease ass  
Human chromosome 1  
USF2a, an inhibito  
Drosophila melanog  
Human protein sequ  
Novel human diagno  
Human protein SEQ  
Human polypeptide  
Human protein SEQ  
Human ischaemic co  
Human protein SEQ  
Human protein sequ  
Human intracellular  
Human prostate can  
Drosophila melanog  
Human transcriptio  
Human polypeptide  
Myosin tail Myosin  
Human polypeptide  
Human polypeptide  
Human PKIN-9 prote  
Human WNK1 protein  
Arabidopsis thalia  
Arabidopsis thalia  
Human adipocyte se  
Human protein sequ  
Prostate cancer-as

FT	Region	75..77
FT	/note= "beta61 strand"	
FT	Region	79..85
FT	/note= "beta7 strand"	
FT	Region	90..92
FT	/note= "beta8 strand"	
FT	Region	103..106
FT	/note= "betas strand"	
FT	Region	119..123
FT	/note= "alpha 2 helix"	
FT	Region	124..130
FT	/note= "alpha 3 helix"	
FT	Region	132..134
FT	/note= "beta10 strand"	
FT	Region	140..142
FT	/note= "beta11 strand"	
XX		
PN	WO200195927-A1.	
PD		
PF	20-DEC-2001.	
PR	14-JUN-2000; 2000WO-KR00630.	
PR	14-JUN-2000; 2000WO-KR00630.	
XX		
PA	(IMAG-) IMAGEGENE CO LTD.	
PI	Kim S, Ko Y, Kim YS, Jo YJ;	
PI	WPI; 2002-096017/13.	
DR	N-PSDB; ABA94640.	
PT	Novel anti-tumour and anti-angiogenic agent of p43 comprises N-terminal domain and C-terminal domain containing eleven beta-strands forming a structural core and three flanking alpha-helices -	
XX		
PS	Claim 1; Fig 1; 35pp; English.	
XX	The invention provides an anti-tumour and anti-angiogenic agent of p43 consisting of two domains, the N-terminal domain (146 amino acids) and C-terminal domain (166 amino acids) containing 11 beta-strands forming a structural core and 3 flanking alpha-helices. p43 is useful as an anti-tumour and anti-angiogenic agent. p43 and its C-terminal cytokine domain (EMAP II) induce regression of fibrosarcoma in immunocompromised mouse while its N-terminal domain does not. p43 is a potent cytokine as determined by the induction of tumour necrosis factor-alpha (TNF-alpha), interleukin-6 (IL-6), IL-8 and matrix metalloproteinase-9 or by its activity of chemotaxis. The present sequence represents the human p43 polypeptide.	
XX		
XX	Sequence 312 AA;	

	Query Match	Similarity	Score	DB	Length
Beet	Local	100.0%	516	23	312
Matches	108	Conservative	0	Mismatches	0
				Indels	0
				Gaps	0

RESULT 2	
AAB63244	
ID	AAB63244 standard; Protein; 328 AA

AC	AAB63244;
XX	
DT	26-MAR-2001 (first entry)

XX	Human breast cancer associated antigen protein sequence SEQ ID NO:606.
XX	
KW	Human; breast cancer; gastric cancer; prostate cancer; diagnosis;
KW	cancer associated antigen; cytostatic; cancer vaccine.
XX	
OS	Homo sapiens.
XX	
PN	WO200073801-A2.
XX	
PD	07-DEC-2000.
XX	
FP	26-MAY-2000; 2000WO-US14749.
XX	
PR	28-MAY-1999; 99US-0136526.
PR	10-SEP-1999; 99US-0153454.
XX	
PA	(LUDW-) LUDWIG INST CANCER RES.
XX	
PI	Obata Y;
XX	
DR	WPI; 2001-025274/03.
XX	
PT	Nucleic acids encoding breast, gastric and prostate cancer associated
PT	antigen precursors, useful for diagnosing and treating a condition
PT	characterized by expression of an abnormal amount of a protein, e.g.
PT	cancer -

PS Example 1, Page 478, 799pp, English.

Seq	Sequence	328 AA;
Query	Match	100.0%; Score 516; DB 22; Length 328;
	Best Local Similarity	100.0%; Pred. No. 1.9e-45;
	Matches 108; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
Qy	1 MANNDAVLKRLKQGGARADQIIEYLKQOVSILKKKATLQATLRKKLRVFNATLKKEIE	60
Db	17 MANNDAVLKRLKQGGARADQIIEYLKQOVSILKKKATLQATLRKKLRVFNATLKKEIE	76
Qy	61 ELKQELVQAEIQNGVKQIAPPSGPTLHANSVSENVIOSTAVTVSSG	108
Db	77 ELKQELVQAEIQNGVKQIAPPSGPTLHANSVSENVIOSTAVTVSSG	124

RESULT 3	
AAR72577	
ID	AAR72577 standard; Protein; 310 AA.
XX	
AC	
XX	AAR72577;
DT	25-MAR-2003 (updated)
DT	29-SEP-1995 (first entry)

XX ENAII; endothelial monocyte activating polypeptide II; chemotaxis  
KW inflammation; tissue factor; tumor; cancer; therapy; metha;  
KN sarcoma.

OS	Mus musculus.
XX	
PN	W09509180-A1.



XX	06-APR-1995.
PD	
XX	29-SEP-1994;
XX	94MO-US11085.
PF	
XX	29-SEP-1993;
PR	93US-0129456.
XX	
PA	(UYCO ) UNIV COLUMBIA NEW YORK.
XX	
PI	Clauss M, Kao J, Kaylon M, Libutti SK, Stern DM;
XX	
DR	WPI, 1995-147389/19.
DR	N-PSDB; AAQ86718.
XX	
PT	New endothelial monocyte activating polypeptide II - induces
PT	chemotaxis, inflammation and tissue factor, useful for treating
PT	tumours, also related antibodies, DNA and active fragments
XX	
PS	Disclosure; Fig.4; 180pp; English.
XX	
CC	A mouse metha sarcoma cDNA library was screened with a probe
CC	based on the N-terminal sequence of mouse EMAP11. Overlapping
CC	clones were combined to obtain a contiguous full-length sequence
CC	(given in AAQ86718) encoding a 33 kDa protein (AAR72577) .
CC	Recombinant EMAP11 was expressed in E. coli.
CC	(Updated on 25-MAR-2003 to correct PN field.)
XX	
Q0	Sequence 310 AA;

Query Match	76.6%;	Score 395;	DB 16;	Length 310;
Best Local Similarity	77.6%;	Pred. No. 7.1e-33;		
Matches 83; Conservative		8; Mismatches 16;	Indels 0;	Gaps 0;

QY 1 MANDVALKRLKLEGGAGADQIIIEILKQOVALLKRAKIILOATLKEKKLRVFNANKKEIE 60

Db 1 MATNDVALKRLKLEGGAGADQIIIEILKQOVALLKRAKIILOATLKEKKLRVFNANKKEIE 60

QY 61 ELKQELIIQAEIÖNGVQIOAFPSGPTLHANSMSVENVIQSTAVTVVSS 107

61 ELKQELIIAEIHNGVEQVRRLSPLOTNCTAASVQSPSVATVTS 107

RESULT 4  
ABB59210  
ID ABB59210 standard; Protein; 294 AA

DT	26-MAR-2002	(first entry)
XX		
DE	Drosophila melanogaster	polypeptide SEQ ID NO 4422.
XX		
KM	Drosophila; developmental biology; cell signalling; insecticide	
KM	pharmaceutical.	
XX		
OS	Drosophila melanogaster.	
XX		
PN	WO200171042-A2.	
XX		
PD	27-SEP-2001.	
XX		
PP	23-MAR-2001; 2001WO-US09231.	
XX		
PR	23-MAR-2000; 2000US-191637P.	
PR	11-JUL-2000; 2000US-0614150.	
XX		
PA	(PEKE ) PE CORP NY.	
XX		
PI	Venter JC, Adams M, Li PWD, Myers EM;	
XX		
DR	WPI; 2001-656860/75.	
DR	N-PSDB; ABLO3313.	
XX		

PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from *Drosophila* and for elucidating cell signalling and cell-cell  
PT interactions -

PS Disclosure; SEQ ID NO 4422; 21pp + Sequence Listing; English  
vvy

CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from *Drosophila*. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (AB161716-AB130511), expressed DNA  
CC sequences (AB01840-AB1616175) and the encoded proteins  
CC (AAB57737-AB012072).

CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://ftp.wipo.int/pub/published_pct_sequences).

**SQ Sequence 294 AA;**

Query Match	20.5%	Score 106;	DB 22;	Length 294;
Best Local Similarity	31.6%;	Pred. No. 0.0089;		
Matches	24;	Conservative	21;	Mismatches 27;
			Indels	4;
			Gaps	1;

```
QY      6 AVLRLEQKGAEDQIIIEYLKQOVSLLEKALIQATLEBKKLRYENAKLKEIELEKQ 65
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      2 ADLQIASNNERRAALINSIEAISGIQ----QLVERQKQELLKENAALAKEVEALAQ 57
```

```
QY      66 LIQAEIQNGVKQIAFP 81
        |:::| | | |
Db      58 LVQLELRNGKKQIPVP 73
```

RESULT 5  
AAB21233  
ID AAB21233 standard; Protein; 672 AA

XX  
DT 09-MAR-2001 (first entry)  
XX  
DB  
XX Corn MFPI.  
XX  
KW Corn; MAR binding filament-like protein 1; MFPI;  
matrix attachment region; MAR; NtMFPI-2; anchor protein

**Zea mays.**

PN W0200061615-A2.

PD 19-OCT-2000.

PF 12-APR-2000; 2000WO-US09723.

PR 12-APR-1999; 99US-0128900.

PA (DUPO) DU PONT DE NEMOURS &amp; CO E I.

PI Harder PA, Meier I;

DR WPI; 2000-679464/66.

XX

PT Nucleic acid fragments from tobacco, corn, soybean and rice, encoding  
PT proteins that are homologs to the MAR binding filament-like protein 1

PT (MFP1), useful for development of novel phenotypes -  
XX  
PS Claim 4; Page 59-61; 62pp; English.

CC The present sequence is a corn matrix attachment region (MAR)  
CC binding filament-like protein 1 (MFPI). A corn EST sequence was  
CC identified as an MFPI homologue from clone p0118 chs4b4r. MFPI has  
CC features of a novel anchor protein that most likely connects chromatin



CC contains a putative wall/membrane anchoring domain. No homology was  
CC found, at the nucleotide and amino acid level, to any other submitted  
CC sequences in existing databases. The FBP binds horse fibrinogen, and  
CC protects against S. equi infection (which causes the disease strangles)  
CC in horses. The protein, or its truncated form, can be used for the  
CC preparation of a vaccine against strangles infection in horses.

XX Sequence 534 AA;

Query Match 18.0%; Score 93; DB 19; Length 534;

Best Local Similarity 29.9%; Pred. No. 0.45; Mismatches 51; Indels 30; Gaps 5;

Matches 41; Conservative 15; Mismatches 51; Indels 30; Gaps 5;

QY 2 ANNDVLRLEQ-----KGADQIIEYLKQVSL--LKEKAILQATLREKRLRVE-- 51

DB 385 AASDAKVALEKEVEAAKAEVADLKQIAKKEEELAEVKEKEALEAKIEBELKTAHAEEL 444

QY 52 -----NAKLKKEIEELKQEL--IOAEIONGVKQIAFPSTGPLHANSM----- 91

DB 445 SKLKEMLERKDHANADLOAEINRLKQELADRIKSLSGGRASQTNPPTTAKAGQLPSTG 504

QY 92 VSENVIOSTAVTVSSG 108

DB 505 ESANPFTIALTVIAG 521

RESULT 8  
AAM98948  
ID AAM98948 standard; Protein; 534 AA.

AC AAM98948;

DT 06-MAY-1999 (first entry)

XX Streptococcus equi M-protein.

XX Streptococcus equi M-protein.

KM Streptococcus equi; M-protein; Sem; immune response; horse; vaccine;

KW abscess; infection.

XX Streptococcus equi.

XX Key Location/Qualifiers

FT Peptide 1..36

FT Protein 37..534

FT /label= Sem

FT /note= "M-like protein"

FT Misc-difference 500..505

FT /note= "membrane anchor sequence"

PN MO9858945-A1.

PD 30-DEC-1998.

PF 23-JUN-1998; 98WO-US12962.

PR 23-JUN-1998; 98US-0103664.

PR 24-JUN-1997; 97US-0050577.

XX (ARTI/) ARTIUSHIN S.

XX (TIMO/) TIMONEY J F.

XX Artushin S, Timoney JF;

XX WPI, 1999-105605/09.

XX N-PSDB; AAX18839.

XX Compounds encoding the protective M-like Streptococcus equi protein

PT - useful for inducing S. equi-specific immunity and identifying S.

XX equi infection in horses

XX Claim 7, Page 6-7, 34pp; English.

CC The present sequence represents Sem, an M-like protein isolated from  
CC Streptococcus equi. A vaccine comprising Sem is useful for inducing  
CC S. equi-specific immunity in a horse. Sem, derived from a horse  
CC biological sample, can be detected in a PCR assay using primers given  
CC in AAX18840 and AAX18841, for identifying horses infected with S. equi.  
CC Compounds from the present invention enable quick assaying of S. equi  
CC infection in horses, which is very important in the early stages of an  
CC outbreak. Vaccines comprising the compounds avoid the risk of reversion  
CC to virulence and occasional abscess development in vaccinated horses.

XX Sequence 534 AA;

Query Match 18.0%; Score 93; DB 20; Length 534;

Best Local Similarity 29.2%; Pred. No. 0.45; Mismatches 51; Indels 30; Gaps 4;

Matches 40; Conservative 16; Mismatches 51; Indels 30; Gaps 4;

QY 2 ANNDVLRLEQKGAADQIIEYLKQV-----SLKKAAILQATLREKRLRVE-- 51

DB 385 AASDAKVALEKEVEAAKAEVADLKQIAKKEEELAEVKEKEALEAKIEBELKTAHAEEL 444

QY 52 -----NAKLKKEIEELKQEL--IOAEIONGVKQIAFPSTGPLHANSM----- 91

DB 445 SKLKEMLERKDHANADLOAEINRLKQELADRIKSLSGGRASQTNPPTTAKAGQLPSTG 504

QY 92 VSENVIOSTAVTVSSG 108

DB 505 ESANPFTIALTVIAG 521

RESULT 9  
AAM78507  
ID AAM78507 standard; Protein; 1047 AA.

AC AAM78507;

DT 06-NOV-2001 (first entry)

XX Human protein SEQ ID NO 1169.

XX Human protein SEQ ID NO 1169.

KM Human, cytokine; cell proliferation; cell differentiation; gene therapy;

KW vaccine; peptide therapy; stem cell growth factor; hematopoiesis;

KW tissue growth factor; immunomodulatory; cancer; leukaemia;

XX nervous system disorder; arthritis; inflammation.

XX Homo sapiens.

XX WO200157190-A2.

XX 09-AUG-2001.

XX 05-FEB-2001; 2001WO-US04098.

XX 03-FEB-2000; 2000US-0496914.

XX 27-APR-2000; 2000US-0560875.

XX 20-JUN-2000; 2000US-0598075.

XX 19-JUL-2000; 2000US-0620325.

XX 01-SEP-2000; 2000US-0654936.

XX 15-SEP-2000; 2000US-0663561.

XX 20-OCT-2000; 2000US-0693125.

XX 30-NOV-2000; 2000US-0728422.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Dmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y,

XX Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;

XX Xue AJ, Yang Y, Wejthman T, Goodrich R;

XX WPI; 2001-476283/51.

XX N-PSDB; AAK51640.

XX Nucleic acids encoding polypeptides with cytokine-like activities,

PT useful in diagnosis and gene therapy -

XX



PA (CURA-) CURAGEN CORP.  
 PI Shinkets RA, Leach M;  
 XX  
 XX WPI; 2000-602362/57.  
 DR N-PSDB; AAC75852.  
 XX  
 PT Novel nucleic acids and peptides derived from open reading frame X,  
 PT useful for treating e.g. cancers, proliferative disorders,  
 PT neurodegenerative disorders and cardiovascular disease -  
 XX  
 PS Claim 11; Page 2052-2053; 5507pp; English.  
 CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
 CC sequences have activities such as: cytostatic; hepatotropic; vulnary;  
 CC antiproliferative; antipartitioning; neurotropic; neuroprotective;  
 CC osteoparitic; anticonvulsant; antiarthritic; immunosuppressant;  
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasodilator;  
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
 CC antiinflammatory; antibacterial; antiviral; antifungal; antineumatic;  
 CC antihypertoid; and antianemic. The sequences can be used for determining  
 CC the presence of or predisposition to, or preventing or treating  
 CC pathological conditions associated with an ORFX-associated disorder. The  
 CC nucleic acids can be used to express ORFX proteins in gene therapy  
 CC vectors. The proteins and nucleic acids may be used to treat cancers,  
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
 CC coagulation; to inhibit thrombosis; and as a contraceptive.  
 XX  
 SQ Sequence 471 AA;  
 Query Match 17.4%; Score 90; DB 21; Length 471;  
 Best Local Similarity 27.0%; Pred. No. 0.78;  
 Matches 30; Conservative 21; Mismatches 48; Indels 12; Gaps 3;  
 QY 8 LKRLQGAADQI-----IEYLKQVSLKEKAILQATLREKKLRVENAKLKEIE 60  
 Db 129 VRELEQANDLIERAKRATIVSLETINKLQAIERNALFESLDERKESILVSGQLKDEAR 188  
 QY 61 ELKQELQATQNGVKQIAPFGTFLANSVSENVIQ---STAVTVSSG 108  
 Db 189 DLROELAVREKQAEVTRKSPSPTLCEKX--DSAVQASISLPATPVGKG 237  
 RESULT 12  
 AAU35717  
 ID AAU35717 standard; Protein; 856 AA.  
 XX  
 AC AAU35717;  
 XX  
 DT 14-FEB-2002 (first entry)  
 XX  
 DE Helicobacter pylori cellular proliferation protein #30.  
 XX  
 KM Antisense; prokaryotic cellular proliferation protein;  
 KM antibiotic; antibacterial; drug design.  
 XX  
 OS Helicobacter pylori.  
 XX  
 PN WO200170955-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 21-MAR-2001; 2001WO-US09180.  
 XX  
 XX 21-MAR-2000; 2000US-191078P.  
 PR 23-MAY-2000; 2000US-206848P.  
 PR

PR 26-MAY-2000; 2000US-207727P.  
 PR 23-OCT-2000; 2000US-242578P.  
 PR 27-NOV-2000; 2000US-253625P.  
 PR 22-DEC-2000; 2000US-257931P.  
 PR 16-FEB-2001; 2001US-263088P.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.  
 PI Haeelbeck R, Ohlsen KL, Zykkind JW, Wall D, Trawick JD, Carr GU;  
 PI Yamamoto RT, Xu HH;  
 XX  
 DR WPI; 2001-611495/70.  
 DR N-PSDB; AAS53576.  
 XX  
 PT New polynucleotides for the identification and development of  
 PT antibiotics, comprise sequences of antisense nucleic acids -  
 XX  
 PS Example 3; Seq ID No 11310; 511pp; English.  
 CC The invention relates to antisense inhibitors of genes essential to  
 CC prokaryotic cellular proliferation, their use in identifying the  
 CC genes, their use in the discovery of novel antibiotics, the essential  
 CC genes themselves and the encoded proteins. The prokaryotes used are  
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
 CC invention is also useful for the identification of potential new targets  
 CC for antibiotic development. The antisense nucleic acids can also be used  
 CC to identify proteins used in proliferation, to express these proteins,  
 CC and to obtain antibodies capable of binding to the expressed proteins.  
 CC The proteins can be used to screen compounds in rational drug discovery  
 CC programmes. The antisense nucleic acid sequence is also useful to screen  
 CC for homologous nucleic acids which are required for cell proliferation in  
 CC a wide variety of organisms. The present sequence represents an  
 CC essential prokaryotic cellular proliferation protein.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 SQ Sequence 856 AA;  
 Query Match 17.2%; Score 89; DB 22; Length 856;  
 Best Local Similarity 38.2%; Pred. No. 2.2;  
 Matches 21; Conservative 14; Mismatches 20; Indels 0; Gaps 0;  
 QY 11 LEOKGAADQIIEYLKQVSLKEKAILQATLREKKLRVENAKLKEIEELKOE 65  
 Db 428 MEKESNAKMOEILKELSDLKEEKIOLKQAFENKEKEVFKEISRLKMEESLAKKE 482  
 RESULT 13  
 AAB96721  
 ID AAB96721 standard; Protein; 1177 AA.  
 XX  
 AC AAB96721;  
 XX  
 DT 29-OCT-2001 (first entry)  
 XX  
 DE Putative P. abyssi ATPase involved in DNA repair #3.  
 XX  
 KM Hyperthermophilic archaeon; hyperthermophilic protein.  
 KM  
 OS Pyrococcus abyssi.  
 XX  
 PN FR2792651-A1.  
 XX  
 PD 27-OCT-2000.  
 XX  
 PF 21-APR-1999; 99FR-0005034.  
 XX  
 XX 21-APR-1999; 99FR-0005034.  
 PR (CNRS ) CNRS CENT NAT RECH SCI.  
 PA

PA (IFRE-) IFREMER INST FR RECH EXPL MER.  
XX  
PI Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;  
PI Querellou J, Weissenbach J, Saurin W, Heilig R;  
XX MPI; 2001-126236/14.  
XX  
PT New nucleotide sequences isolated from *Pyrococcus abyssi* encode  
PT proteins useful in industry -  
XX  
PS Claim 7; Pages 1483-1487, 1657pp; French.  
XX  
CC The present invention relates to the genomic sequence of *Pyrococcus*  
CC *abyssi* (see AAB66431 and AAB41223-7) and *P. abyssi* proteins. *P. abyssi* is  
CC a hyperthermophilic archaeon, which is isolated from deep-sea  
CC hydrothermal vents. The present sequence is one such *P. abyssi* protein.  
CC The proteins of the present invention have various potential industrial  
CC uses, since the proteins are stable at very high temperatures, some up to  
CC 110 degrees centigrade.  
CC Note: This patent is in the same patent family as WO2000065062, which  
CC contains additional sequences as shown in AAB99132-AAB99143,  
CC AAB75903-AAB75920 and AAB66436.  
XX  
SQ Sequence 1177 AA;  
  
Query Match 17.2%; Score 89; DB 22; Length 1177;  
Best Local Similarity 29.3%; Pred. No. 3.4; Indels 26; Gaps 3;  
Matches 23; Conservative 20; Mismatches 24;  
  
QY 4 NDAVLKLEQKGAADQIIIEYLK-----QVSLKEX-----ALLQA 40  
DB 408 NEADIKRLAEKERLSRITLAKKLGIREVEKLEKLEKKAELSVENKTSISQR 467  
QY 41 TLREKTLRVENAKLTK--EIEELKQELIQAEIQNVK 76  
DB 468 RRKVEELERKTSSELQKSSLSLSELERELIKAEQSEVR 506  
  
RESULT 14  
ABG06864  
ID ABG06864 standard; Protein; 472 AA.  
XX  
AC ABG06864;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #6855.  
XX  
KM Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KM food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR MPI; 2001-639362/73.  
DR N-PSDB; AAS71051.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -

XX  
PS Claim 20; SEQ ID No 37223; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 472 AA;  
  
Query Match 17.2%; Score 88.5; DB 22; Length 472;  
Best Local Similarity 26.9%; Pred. No. 1.1; Indels 7; Gaps 2;  
Matches 25; Conservative 29; Mismatches 32;  
  
QY 8 LKRLKQKGAADQIIIEYLKQVSLKERAIIQATLREKLRVENAKLKEIEELKQELI 67  
DB 383 LKRLDVEYVGQAIFLEFLKEQISLAEKKML-----DYRLRLRAEQKQVEESLRKLL 436  
QY 68 QAEIQ-NGYKQIAFPGSGPLHANSVSENVIOS 99  
DB 437 VAEINRLQAVEALCSSQHTMIESNDSEETIRT 469  
  
RESULT 15  
AAO17717  
ID AAO17717 standard; Protein; 667 AA.  
XX  
AC AAO17717;  
XX  
DT 08-AUG-2002 (first entry)  
XX  
DE O cuniculus endooligopeptidase A related protein.  
XX  
KM Human; protease associated protein; HPRAP; cancer; inflammation;  
KM anti-HIV; antinaemic; antiarteriosclerotic; antiasthmatic; cytostatic;  
KM antiinflammatory; hepatotropic; antidiabetic; nephrotropic; antitumor;  
KM antihydroid; immunosuppressive; thyromimetic; virucide; dermatological;  
KM neuroprotective; cardiac; osteopathic; antidiabetic; antiparasitic;  
KM antiparasitic; uropachic; ophthalmological; antirheumatic; tranquilliser;  
KM vulnerary; antitumor; immune disorder; gene therapy; HPRAP-1.  
XX  
OS *Oryctolagus cuniculus*.  
XX  
PN WO9957274-A1.  
XX  
PD 11-NOV-1999.  
XX  
PF 28-APR-1999; 99WO-US09190.  
XX  
PR 01-MAY-1998; 98US-0071709.  
XX  
PA (INCY-) INCYTE PHARM INC.  
XX  
PI Hillman JL, Tang YT, Lal P, Corley NC, Guegler KJ, Patterson C;  
XX MPI; 2000-062147/05.  
DR

XX Novel human protease associated proteins used for, e.g. the diagnosis  
PT and prevention of cell proliferative and immune disorders  
XX  
PS  
XX Disclosure; Page -; 75pp; English.

CC The present invention provides the protein and coding sequences of four  
CC human protease associated proteins, designated HPRAP-1, HPRAP-2, HPRAP-3  
CC and HPRAP-4. Human protease associated proteins are expressed in cancer  
CC and immortalised cell lines and tissues associated with inflammation and  
CC the immune response, and so appear to play a role in cell proliferative  
CC and immune disorders. The sequences can be used to diagnose, treat or  
CC prevent cell proliferative and immune disorders, including actinic  
CC keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis,  
CC hepatitis, mixed connective tissue disease, myelofibrosis, paroxysmal  
CC nocturnal haemoglobinuria, polycythaemia vera, psoriasis, primary  
CC thrombocythaemia, AIDS, Addison's disease, adult respiratory distress  
CC syndrome, allergies, ankylosing spondylitis, amyloidosis, anaemia,  
CC asthma, atherosclerosis, autoimmune haemolytic anaemia, autoimmune  
CC thyroiditis, bronchitis, cholecystitis, contact dermatitis, Crohn's  
CC disease, atopic dermatitis, dermatomyositis, diabetes mellitus,  
CC emphysema, episodic lymphopenia with lymphocytotoxicity, erythroblastosis  
CC fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis,  
CC Goodpasture's syndrome, gout, Grave's disease, Hashimoto's thyroiditis,  
CC hyperostosis, irritable bowel syndrome, multiple sclerosis,  
CC myasthenia gravis, myocardial or pericardial inflammation,  
CC osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis,  
CC Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjogren's syndrome,  
CC systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis,  
CC thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome,  
CC infections and trauma. The present sequence is the rabbit  
CC endooligopeptidase A related protein, which was used in a homology  
CC comparison in the exemplification of the invention.  
CC Note: The present sequence was not obtainable from this Derwent basic but  
CC was obtained from the Derwent family equivalent US6171790.

XX  
SQ Sequence 667 AA;

Query Match 17.1%; Score 88; DB 21; Length 667;

Best Local Similarity 29.1%; Pred. No. 2;  
Matches 32; Conservative 18; Mismatches 50; Indels 10; Gaps 3;

QY 2 ANNDVLRLEQKGAADQIIIEYLKQVSLKKAIIQATLRBEKKRVENAKLKEIE 61  
DB 332 ANDD-----LERAKRATIVSLTETKLNQAIERNAFLESELDKESLVSQRLDEARD 386  
QY 62 LKQELIQAEIQNGVQKQIAFPSCGTPHANSVMSENVIQ---STAVTVSSG 108  
DB 387 LROELAVREROQEVTRKSPSPPTLDCEKM--DSAVQASLSLPATPVGKG 434

Search completed: January 2, 2004, 18:20:31  
Job time : 33.1882 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using SW model

Run on: January 2, 2004, 18:18:32 ; Search time 11.8588 seconds  
(without alignments)  
875.822 Million cell updates/sec

Title: US-09-930-169-2

Perfect score: 516  
Sequence: 1 MANDAVIKRLKQKGAEDQ.....NSMSENVIOSTAVTTVSSG 108

Scoring table: BLOSUM62  
Gapop: 10.0 , Gapext: 0.5

Searched: 283308 segs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	511	99.0	310	2 B55053	endothelial monocy
2	395	76.6	310	2 A55053	endothelial monocy
3	101	19.6	270	2 G71061	hypothetical prote
4	96	18.6	163	2 D71169	hypothetical prote
5	94	18.2	213	2 B83972	hypothetical prote
6	93.5	18.1	662	2 A54078	methyl-accepting c
7	93	18.0	1130	2 T34081	hypothetical prote
8	92	17.8	1169	2 A64505	p115 homolog - Met
9	91.5	17.7	507	2 S05542	hypothetical prote
10	90.5	17.5	162	2 A56254	nuclear b2ip prote
11	89.5	17.3	843	2 S44868	kinesin heavy chai
12	89	17.2	160	2 H75062	flagella-related p
13	89	17.2	856	2 H64552	endopeptidase Clp
14	89	17.2	1177	2 B75150	chromosome segrega
15	88	17.1	212	2 S42396	USF2 protein - mou
16	88	17.1	346	2 A55111	transcription fact
17	88	17.1	346	2 J67638	upstream stimulat
18	88	17.1	512	2 PC7064	endoilgopeptidase
19	87.5	17.0	1079	2 T38913	translational initia
20	87.5	17.0	2116	2 A26653	myosin heavy chain
21	87	16.9	234	2 I54074	upstream stimulat
22	86.5	16.8	389	2 B44972	paramyosin - nemat
23	86	16.7	955	2 S24348	myosin heavy chain
24	86	16.7	1024	2 T30868	Rho-binding prote
25	86	16.7	1256	2 T26101	hypothetical prote
26	86	16.7	1938	1 JX0178	myosin heavy chain
27	85.5	16.6	638	2 S36723	FUN36 protein - ye
28	85	16.5	348	2 C70417	phosphate starvati
29	85	16.5	369	2 I78877	CAMP responsive el

30	85	16.5	508	2 A45477	CAMP response elem
31	85	16.5	752	2 A65089	DNA topoisomerase
32	85	16.5	752	2 G91116	DNA topoisomerase
33	85	16.5	752	2 G85961	DNA topoisomerase
34	85	16.5	761	2 T43788	DNA topoisomerase
35	84.5	16.4	375	2 T12251	hypothetical prote
36	84.5	16.4	662	2 D54078	hypothetical prote
37	84.5	16.4	662	2 D82814	ATP-dependent Clp
38	84.5	16.4	1046	2 T42720	cytoplasmic linker
39	84.5	16.4	1046	2 T42734	cytoplasmic linker
40	84.5	16.4	1526	2 T41522	myosin ii - fisio
41	84.5	16.4	2288	2 T29959	hypothetical prote
42	84	16.3	400	2 E70318	hypothetical prote
43	84	16.3	752	2 A45582	DNA topoisomerase
44	84	16.3	1937	2 I38055	myosin heavy chain
45	83.5	16.2	394	2 T03998	ATG1 protein homol

## ALIGNMENTS

```

RESULT 1
B55053
endothelial monocyte-activating protein II precursor - human
C/Species: Homo sapiens (man)
C/Date: 08-Jul-1995 #sequence_revision 03-Aug-1995 #text_change 09-Dec-2002
C/Accession: B55053
R/Kao, J.; Houck, K.; Fan, Y.; Haehnel, I.; Libutti, S.K.; Kayton, M.L.; Grikscheit, T.;
J., J.; Stern, D.M.
J. Biol. Chem. 269, 25106-25119, 1994
A/Title: Characterization of a novel tumor-derived cytokine. Endothelial-monocyte active
A/Reference number: A55053; MUID:95014290; PMID:7929199
A/Accession: B55053
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-310 <KAO>
A/Cross-references: GB:U0117
C/Suprafamily: endothelial monocyte-activating polypeptide II precursor (pro-EMAP II)

Query Match          99.0%; Score 511; DB 2; Length 310;
Best Local Similarity 99.1%; Pred. No. 5.2e-31;
Matches 107; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MANDAVIKRLKQKGAEDQIIETLKQVSLKKEKATLREKKLRVENAKLKEIE 60
DB 1 MANDAVIKRLKQKGAEDQIIETLKQVSLKKEKATLREKKLRVENAKLKEIE 60
QY 61 ELKQELIQAEIQNGVKQIAPSGTPLHANSWSENVIOSTAVTTVSSG 108
DB 61 ELKQELIQAEIQNGVKQIAPSGTPLHANSWSENVIOSTAVTTVSSG 108

RESULT 2
A55053
endothelial monocyte-activating protein II precursor - mouse
C/Species: Mus musculus (house mouse)
C/Date: 08-Jul-1995 #sequence_revision 03-Aug-1995 #text_change 09-Dec-2002
C/Accession: A55053; MUID:95014290; PMID:7929199
R/Kao, J.; Houck, K.; Fan, Y.; Haehnel, I.; Libutti, S.K.; Kayton, M.L.; Grikscheit, T.;
J., J.; Stern, D.M.
J. Biol. Chem. 269, 25106-25119, 1994
A/Title: Characterization of a novel tumor-derived cytokine. Endothelial-monocyte active
A/Reference number: A55053; MUID:95014290; PMID:7929199
A/Accession: A55053
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-310 <KAO>
A/Cross-references: GB:U0118; NID:g498911; PID:AAA62203.1; PID:g498912
R/Kao, J.; Ryan, J.; Bretz, G.; Chen, J.; Shen, H.; Fan, Y.G.; Godman, G.; Familletti, P
J. Biol. Chem. 267, 20239-20247, 1992
A/Title: Endothelial monocyte-activating polypeptide II. A novel tumor-derived polypeptide
A/Reference number: A44032; MUID:93015897; PMID:1400342
A/Accession: A44032

```

A:Molecule type: protein  
A:Residues: 145-158, 'X', 160-164 <KA2>  
A:Experimental source: methylcholanthrene A fibrosarcoma cells  
A:Note: sequence extracted from NCBI backbone (NCBI:115676)  
C:superfamily: endothelial monocyte-activating polypeptide II precursor (pro-EMAP II)

Query Match 76.6%; Score 395; DB 2; Length 310;  
Best Local Similarity 77.6%; Pred. No. 2.1e-22;  
Matches 83; Conservative 8; Mismatches 16; Indels 0; Gaps 0;

QY 1 MANDAVLRLEKQKGAADIIIEYLKQVSLKEKAILQATLREKKLRVENAKLKEKE 60  
DB 1 MANDAVLRLEKQKGAADIIIEYLKQVSLKEKAILQATLREKKLRVENAKLKEKE 60

QY 61 ELKQELIOAEIONGVKQIAFPSPGTPHANSVSENYIOSTAVTVSS 107  
DB 61 ELKQELIOAEIONGVKQIAFPSPGTPHANSVSENYIOSTAVTVSS 107

RESULT 3  
hypothetical protein PH189 - Pyrococcus horikoshii  
C:Species: Pyrococcus horikoshii  
C>Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 05-Nov-1999

C/Accession: G71061  
R:Kawarayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, S.; Sekin  
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kusida, N.; Oguchi  
DNA Res. 5, 55-76, 1998  
A>Title: Complete sequence and gene organization of the genome of a hyper-thermophilic  
A:Reference number: A71000; MUID:98344137; PMID:9679194  
A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-270 <KAW>

A/Cross-references: GB:AP000005; NID:g3236132; PIDN:BA30289.1; PID:d1031232; PID:g32576  
A:Experimental source: strain OT3  
A:Note: this accession replaces an interim accession for a sequence replaced by GenBank  
C:Genetics:  
A:Gene: PH189

Query Match 19.6%; Score 101; DB 2; Length 270;  
Best Local Similarity 30.0%; Pred. No. 1.2;  
Matches 30; Conservative 22; Mismatches 32; Indels 16; Gaps 3;

QY 3 NNDVAVLRLEKQKGAADIIIEYLKQVSLKEKAILQATLREKKLRVENAKL 55  
DB 86 NNDVAVLRLEKQKGAADIIIEYLKQVSLKEKAILQATLREKKLRVENAKL 55

QY 56 KKEIELEKQELIOAEIONGVKQIAFPSPGTPHANSVSENYIOSTAVTVSS 95  
DB 144 KKEIELEKQELIOAEIONGVKQIAFPSPGTPHANSVSENYIOSTAVTVSS 95

RESULT 4  
D71169  
hypothetical protein PH0552 - Pyrococcus horikoshii  
C:Species: Pyrococcus horikoshii  
C>Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 20-Jun-2000

C/Accession: D71169  
R:Kawarayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, S.; Sekin  
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kusida, N.; Oguchi  
DNA Res. 5, 55-76, 1998  
A>Title: Complete sequence and gene organization of the genome of a hyper-thermophilic  
A:Reference number: A71000; MUID:98344137; PMID:9679194  
A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-163 <KAW>

A/Cross-references: GB:AP000002; NID:g3236129; PIDN:BA29641.1; PID:g3256958  
A:Experimental source: strain OT3  
A:Note: this accession replaces an interim accession for a sequence replaced by GenBank  
C:Genetics:  
A:Gene: PH0552

Query Match 18.6%; Score 96; DB 2; Length 163;  
Best Local Similarity 33.3%; Pred. No. 1.7;  
Matches 24; Conservative 19; Mismatches 21; Indels 8; Gaps 2;

QY 8 LKRLKQKGAADIIIEYLKQVSLKEKAILQATLREKKLRVENAKLKEKEBELKQEL 67  
DB 83 LKRLKQKGAADIIIEYLKQVSLKEKAILQATLREKKLRVENAKLKEKEBELKQEL 67

QY 68 OAEIONGVKQIA 79  
DB 136 OAEIONGVKQIA 79

RESULT 5  
E83972  
hypothetical protein BH2581 [imported] - Bacillus halodurans (strain C-125)  
C:Species: Bacillus halodurans  
C>Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001

C/Accession: E83972  
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Maui, N.; Fuji, F.; Hiran  
Nucleic Acids Res. 28, 4317-4331, 2000  
A>Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and  
A:Reference number: A83650; MUID:20512582; PMID:11058132  
A/Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-213 <STO>

A/Cross-references: GB:AP001516; GB:BA000004; NID:g10175192; PIDN:BA06300.1; GSPDB:GN001  
A:Experimental source: strain C-125  
C:Genetics:  
A:Gene: BH2581

Query Match 18.2%; Score 94; DB 2; Length 213;  
Best Local Similarity 36.8%; Pred. No. 3.2;  
Matches 25; Conservative 16; Mismatches 15; Indels 12; Gaps 3;

QY 7 VLKRLKQKGAADIIIEYLKQVSLKEKAILQATLREKKLRVENAKLKEKE 57  
DB 115 VLKRLKQKGAADIIIEYLKQVSLKEKAILQATLREKKLRVENAKLKEKE 57

QY 58 EIELEKQEL 65  
DB 172 EIELEKQEL 65

RESULT 6  
A54078  
methyl-accepting chemotaxis protein mcpB - Bacillus subtilis  
C:Species: Bacillus subtilis  
C>Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 20-Jun-2000

C/Accession: A54078; H69655  
R:Hanson, D.W.; Ordal, G.W.  
J. Biol. Chem. 269, 14038-14046, 1994  
A>Title: Cloning and characterization of genes encoding methyl-accepting chemotaxis prote  
A:Reference number: A54078; MUID:94245722; PMID:8188684  
A/Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-662 <HAN>

A/Cross-references: GB:U129189; NID:g459687  
A:Note: the sequence in GenBank entry BACMCPBP, release 106.0, (PID:g459688) has 360-N  
R:Kunitz, F.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Allison, G.; Azevedo, V.; Berer  
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi  
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Birtington, J.; Fabre, C.; Ferrati, E.  
Nature 350, 249-256, 1997  
A/Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funa, S.; Gallizzi, A.; Gall  
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.;  
Koester, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,  
A.; Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Maeda, S.; Mausel,  
Y.; M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle,  
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon,  
A/Authors: Schleif, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror,

Akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpetra, P.; Tognoni, A.; Tosato, V.; Uchiyama, T.; Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.  
A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.  
A:Reference number: A69580; MUID:98044033; PMID:9384377  
A:Accession: H69555  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-359, 'N', 361-449, 'G', 451-662 <KUN>  
A:Cross-references: GB:299119; GB:299120; GB:AL009126; NID:g2635613; PIDN:CA15115.1; PI  
A:Experimental source: strain 168  
C:Comment: Strains in which this protein is inactivated are defective in chemotaxis towa  
C:Genetics:  
A:Gene: mcpB  
C:Superfamily: Probable methyl-accepting chemotaxis transducer  
C:Keywords: chemotaxis; signal transduction; transmembrane protein

Query Match 18.1%; Score 93.5; DB 2; Length 662;  
Best Local Similarity 29.8%; Pred. No. 12;  
Matches 34; Conservative 16; Mismatches 37; Indels 27; Gaps 4;

Oy 5 DAVIKRLKQKAEADQIIEYLK---QVSLKEKALIA-----TLREKRL 49  
Db 464 EAVVAGLEKSKDITSLRVINGIADQTNLALNLAIPADAGSGSGSVAAERKLA 523

Oy 50 VENAKKEIEELKQELIQAEIONGVQKIAFPSTPLHANSVSENVISTAVT 103  
Db 524 VQSDASAEIEKLEIQLDEIV-ALID-----TSLMFREVNQEVQSGLVVT 565

RESULT 7  
T34081  
hypothetical protein C02F12.7 - *Caenorhabditis elegans*  
C:Species: *Caenorhabditis elegans*  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T34081  
R:Miller, N.  
A:Description: The sequence of *C. elegans* coamid C02F12.  
A:Reference number: Z21473  
A:Accession: T34081  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1130 <ML>  
A:Cross-references: EMBL:U41545; PIDN:AA83190.1; CESP:C02F12.7  
C:Genetics:  
A:Gene: C02F12.7  
A:Introns: 11/2; 118/1; 156/2; 239/2; 308/2; 398/1; 458/3; 591/1; 640/3; 756/2; 790/2; 8

Query Match 18.0%; Score 93; DB 2; Length 1130;  
Best Local Similarity 24.6%; Pred. No. 23;  
Matches 29; Conservative 30; Mismatches 33; Indels 26; Gaps 4;

Oy 2 ANNDVILKRLKQKAEADQIIEYLKQVSLKEKALIAOTLREK-----KLRYENAKL 55  
Db 691 SDKDMLLELSK---NKNIEHLKQIAQNEKISTREKDSLEKTIQLEIDNSK 746

Oy 56 KKEIEEL-----KQELIQAEIONGVQKIAFPSTPLHANSVSENVISTAVT 99  
Db 747 SDQIEKILRVNMDLQWGTIKDELVKNEE--IKTISAKTQALLESNTVSESTYAS 802

RESULT 8  
A64505  
P15 homolog - *Methanococcus jannaschii*  
C:Species: *Methanococcus jannaschii*  
C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 02-Jun-2000  
C:Accession: A64505  
R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, R.; Reich, C.T.; Overbeek, R.; Kirkness, E.F.; Weissbrock, K.G.; Weirick, J.M.; Glodek, A.; Iwou, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.  
Science 273, 1058-1073, 1996  
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C

A:Title: Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*  
A:Reference number: A64300; MUID:96337999; PMID:8688087  
A:Accession: A64505  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-1169 <BUL>  
A:Cross-references: GB:U67604; GB:L77117; NID:g1592224; PID:g1500543; TIGR:MJ1643  
C:Genetics:  
A:Map position: FOR1623481-1626990  
C:Superfamily: chromosome segregation protein SMC1

Query Match 17.8%; Score 92; DB 2; Length 1169;  
Best Local Similarity 26.6%; Pred. No. 28;  
Matches 25; Conservative 24; Mismatches 25; Indels 20; Gaps 4;

Oy 3 NNDVILKRLKQKAEADQIIEYLKQ---QVSLKEKALIAOTLREKRLRYENAKLKE 58  
Db 281 NN--IINELNEKNE--EVLHLKSIKELEVEIENDKVDLSINELKVEVEIENKKE 336

Oy 59 IEELKQELIQ-----AEIONGVQKIAF 80  
Db 337 IKETQKILNRDSITKEQKIEIKKINLVNT 370

RESULT 9  
S05542  
hypothetical protein, 54K - *Enterococcus faecium*  
C:Species: *Enterococcus faecium*  
C:Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 15-Oct-1999  
C:Accession: S05542  
R:Fuerst, P.; Moesch, H.U.; Solioz, M.  
Nucleic Acids Res. 17, 6724, 1989  
A:Title: A protein of unusual composition from *Enterococcus faecium*.  
A:Reference number: S05542; MUID:8938598; PMID:2780257  
A:Accession: S05542  
A:Molecule type: DNA  
A:Residues: 1-507 <FUE>  
A:Cross-references: GB:X16421; EMBL:M26048; NID:g43333; PIDN:CAA4442.1; PID:g43334  
A:Note: the authors translated the codon CGT for residues 221 and 223 as Lys

Query Match 17.7%; Score 91.5; DB 2; Length 507;  
Best Local Similarity 30.3%; Pred. No. 12;  
Matches 33; Conservative 28; Mismatches 37; Indels 11; Gaps 4;

Oy 5 DAVIKRLKQKAEADQIIEYLKQVSLKEKALIAOTLREKRLRYENAKLKEIEIKQ 64  
Db 29 DKTIADIQNOQASQSQIEALEGVSAINTKA--QDLTKQDTLRKESAOIKQIKLOE 86

Oy 65 --ELIQAEIONGVQKIAFPSTPLHANSVSENVISTAVT 105  
Db 87 RIEKREATIQKARETOVKNTSSNYIDAVLNADSL-ADAVGRIOAMSTI 134

RESULT 10  
A56254  
nuclear bzip protein MatG - chicken  
N:Alternate names: Mat oncogene homolog MatG  
C:Species: *Gallus gallus* (chicken)  
C:Date: 05-May-1995 #sequence\_revision 05-May-1995 #text\_change 21-Jul-2000  
C:Accession: I50378; A56254  
R:Kataoka, K.; Igatahshi, K.; Itoh, K.; Fujiwara, K.T.; Noda, M.; Yamamoto, M.; Nishizawa M.; Mol. Cell. Biol. 15, 2180-2190, 1995  
A:Title: Small Mat proteins heterodimerize with Fos and may act as competitive repressors.  
A:Reference number: I50378; MUID:95198738; PMID:7891713  
A:Accession: I50378  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-162 <KAT>  
A:Cross-references: GB:D28602; NID:g516725; PIDN:BA05939.1; PID:g516726  
A:Note: submitted to DDBJ, February 1994  
C:Genetics:  
A:Gene: matG  
A:Introns: 12/3

C/Superfamily: maf homology  
C/Keywords: nucleus; transcription factor  
F:25-114/Domain: maf homology <MAF>

Query Match 17.5%; Score 90.5; DB 2; Length 162;  
Best Local Similarity 27.6%; Pred. No. 4.3;  
Matches 29; Conservative 23; Mismatches 42; Indels 11; Gaps 3;

QY 9 KRLEOKGAEADQIIEYKQVSLKEKAILQATLREKKLRVENAKLKKIEELKQELI 68  
DB 57 RILKMGVYASCKVKTQCELEKQVLEQ---QVEKLASNSAKMELDLRSK--Y 111  
QY 69 AEIQNGVKQIAPSPGTPPLHA-----NSMVSENVIOSTAVTTVS 107  
DB 112 EALQNFARTVARSPTVPVRGRLNSMGPLVPGVATTSVITTKS 156

## RESULT 11

844868  
Kinesin heavy chain unc-116 - Caenorhabditis elegans

N/Alternate names: R05D3.7 protein

C/Species: Caenorhabditis elegans

C/Date: 14-Sep-1994 #sequence\_revision 12-May-1995 #text\_change 23-Mar-2001

C/Accession: S44868; A48264

R/Wilson, R.

submitted to the EMBL Data Library, December 1992

A/Description: Sequence of the C. elegans cosmid R05D3.

A/Reference number: S44733

A/Accession: S44868

A/Molecule type: DNA

A/Residues: 1-843 <NTL>

A/Cross-references: EMBL:L07144; NID:G156412; PIDN:AAA28137.1; PID:G156420

R/Patel, N.; Thierry-Mieg, D.; Mancillas, J.R.

Proc. Natl. Acad. Sci. U.S.A. 90, 9181-9185, 1993

A/Title: Cloning by insertional mutagenesis of a cDNA encoding Caenorhabditis elegans ki

A/Reference number: A48264; MUID:94022338; PMID:8105472

A/Accession: A48264

A/Status: not compared with conceptual translation

A/Molecule type: mRNA

A/Residues: 1-814, 'A' <PAT>

A/Cross-references: GB:U19120; NID:G439589; PIDN:AAA28155.1; PID:G439590

A/Experimental source: mixed-stage hermaphrodite library

A/Note: sequence extracted from NCBI backbone (NCBIP:138249)

C/Genetics:

A/Insertions: 13/3; 45/3; 99/3; 199/1; 275/3; 717/3

C/Superfamily: unassigned kinesin-related proteins; kinesin motor domain homology

C/Keywords: ATP; nucleotide binding; P-loop

F:12-35/Domain: kinesin motor domain homology <KMOT>

F:88-95/Region: nucleotide-binding motif A (P-loop)

Query Match 17.3%; Score 89.5; DB 2; Length 843;  
Best Local Similarity 36.0%; Pred. No. 30;  
Matches 27; Conservative 22; Mismatches 17; Indels 9; Gaps 5;

QY 9 KRLEOKGAEADQIIEYKQVSLKEKAILQATLRE-EKKLRVENAKLKKIEELKQ 64  
DB 440 QQDERDDDEQKXSGELKRLQV-LIQEAL--GTRKREBELIRENNRFGKEADPKQ 496

QY 65 ELIQAEIQNGVKQIA 79  
DB 497 E--GKEWMTALEEIA 509

## RESULT 12

H75062

flagella-related protein C PAB1381 - Pyrococcus abyssi (strain Orsay)

C/Species: Pyrococcus abyssi

C/Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Jun-2000

C/Accession: H75062

R/anonymous, Genoscope

submitted to the EMBL Data Library, July 1999

A/Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome stru

A/Reference number: A75001

A/Accession: H75062  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-160 <KAW>  
A/Cross-references: GB:AJ248287; GB:AL096836; NID:G5458657; PIDN:CA850397.1; PID:G545881C  
A/Experimental source: strain Orsay  
C/Genetics:  
A/Genes: PAB1381

Query Match 17.2%; Score 89; DB 2; Length 160;  
Best Local Similarity 30.6%; Pred. No. 5.4;  
Matches 22; Conservative 21; Mismatches 21; Indels 8; Gaps 2;

QY 8 KRLEOKGAEADQIIEYKQVSLKEKAILQATLREKKLRVENAKLKKIEELKQELI 67  
DB 80 IERLDKTIQDMMLEKVEVSGEINPFEQ-----LSQSSLSFVGEELKQIEELRML- 132

QY 68 QAEIQNGVKQIA 79  
DB 133 -AQIKNDIKVLA 143

## RESULT 13

H64552

endopeptidase Clp ATP-binding chain B - Helicobacter pylori (strain 26695)

N/Alternate names: ATP-dependent Clp proteinase regulatory chain

N/Contains: adenosinetriphosphatase (EC 3.6.1.3)

C/Species: Helicobacter pylori

C/Date: 09-Aug-1997 #sequence\_revision 15-Aug-1997 #text\_change 19-Jan-2001

C/Accession: H64552

R/Tomb, J.F.; White, O.; Kervatage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.;

Peterson, S.; Loftus, B.; Richardson, D.; Dodson, J.; Khalak, H.G.; Glodek, A.; McKenney

son, J.D.; Kelley, J.M.; Cotton, M.D.; Weldman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.

Nature 388, 539-547, 1997

A/Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Frazer, C.A

A/Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.

A/Reference number: A64520; MUID:97394467; PMID:9252185

A/Accession: H64552

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-856 <TOM>

A/Cross-references: GB:AE000545; GB:AE000511; NID:G2313349; PIDN:AA07330.1; PID:G231335C

C/Function:

A/Description: allows clpP to hydrolyze polypeptides and proteins, probably by a chaperon

e activity; ATP hydrolysis is required for Clp hydrolysis of proteins but not of smaller

C/Superfamily: endopeptidase Clp ATP-binding chain

C/Keywords: ATP; duplication; hydrolase; molecular chaperone; nucleotide binding; P-loop

F:204-211/Region: nucleotide-binding motif A (P-loop)

F:271-276/Region: nucleotide-binding motif B

F:605-612/Region: nucleotide-binding motif A (P-loop)

F:673-678/Region: nucleotide-binding motif B

F:210/Binding site: ATP (lys) #status predicted

F:611/Binding site: ATP (lys) #status predicted

Query Match 17.2%; Score 89; DB 2; Length 856;  
Best Local Similarity 38.2%; Pred. No. 34;  
Matches 21; Conservative 14; Mismatches 20; Indels 0; Gaps 0;

QY 11 LEQKGAEDQIIEYKQVSLKEKAILQATLREKKLRVENAKLKKIEELKQ 65  
DB 428 MEKESNAKRMQBIILKELSDLEKEKIQLEAQFENEKEVKEISRLKMEWESLKE 482

## RESULT 14

B75150

chromosome segregation protein (smc1) PAB2109 - Pyrococcus abyssi (strain Orsay)

C/Species: Pyrococcus abyssi

C/Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Jun-2000

C/Accession: B75150

R/anonymous, Genoscope

submitted to the EMBL Data Library, July 1999

A/Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome stru

A/Reference number: A75001



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2004, 18:17:12 ; Search time 8.47059 Seconds  
(without alignments)  
599.590 Million cell updates/sec

Title: US-09-930-169-2

Perfect score: 516

Sequence: 1 NANNDAVVKRLKRGKGAEDQ.....NSMVSENVIOSTAVTVSSG 108

Scoring table: BLOSUM62

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	516	100.0	312	1	MCAL_HUMAN
2	425	82.4	359	1	MCAL_CRICK
3	395	76.6	310	1	MCAL_MOUSE
4	93.5	18.1	662	1	MCBP_BACU
5	93	18.0	1130	1	Y147_CAEEL
6	92	17.8	1169	1	SMC_METJA
7	91.5	17.7	162	1	MARG_HUMAN
8	91.5	17.7	516	1	P54_ENTFC
9	90.5	17.5	162	1	MARG_CHICK
10	90	17.4	956	1	KF5C_MOUSE
11	89.5	17.3	162	1	MARG_MOUSE
12	89.5	17.3	162	1	MARG_MOUSE
13	89.5	17.3	162	1	MARG_MOUSE
14	89	17.2	856	1	KLIN_CAEEL
15	88	17.1	291	1	USF2_RAT
16	88	17.1	346	1	USF2_MOUSE
17	87.5	17.0	1079	1	IF2P_SCHPO
18	87.5	17.0	2116	1	MYS2_DICOT
19	87	16.9	346	1	USF2_HUMAN
20	86	16.7	1024	1	RIP3_MOUSE
21	86	16.7	1029	1	RIP3_MOUSE
22	86	16.7	1938	1	MYS2_CHICK
23	85.5	16.6	638	1	YAB1_YEAST
24	85	16.5	570	1	STRM_DROME
25	85	16.5	752	1	PARC_ECOLI
26	85	16.5	1941	1	MYH2_HUMAN
27	84.5	16.4	736	1	TLPB_BACU
28	84.5	16.4	736	1	MYH7_RABIT
29	84.5	16.4	1526	1	MYS2_SCHPO
30	84	16.3	752	1	PARC_SALTY
31	84	16.3	1937	1	MYH8_HUMAN
32	84	16.3	1939	1	MYH1_HUMAN
33	83.5	16.2	1934	1	MYH7_MESAU

34	83	16.1	171	1	ATPF_HELPY
35	83	16.1	452	1	TIG_CAUCR
36	83	16.1	724	1	HMWR_HUMAN
37	83	16.1	1013	1	AGOD_DROME
38	83	16.1	1093	1	TRF1_HUMAN
39	83	16.1	1690	1	C190_DROME
40	83	16.1	1939	1	MYH4_HUMAN
41	83	16.1	1940	1	MYH3_CHICK
42	82.5	16.0	790	1	MUS2_THETN
43	82.5	16.0	794	1	HMWR_MOUSE
44	82.5	16.0	1084	1	MYS2_RABIT
45	82.5	16.0	1935	1	MYH7_HUMAN

## ALIGNMENTS

RESULT 1  
MCAL\_HUMAN STANDARD; PRT; 312 AA.  
AC Q12904; Q96CQ9; 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Multisynthetase complex auxiliary component p43 [Contains:  
DE Endothelial-monocyte activating polypeptide II (EMAP-II) (Small  
DE Inducible cytokine subfamily E member 1)].  
CN SCYB1 OR EMB2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. PubMed=7929199;  
RX MEDLINE=95014290; Kao J., Houck K., Fan Y., Haehnel I., Libutti S.K., Kayton M.L.,  
RA Grikscheit T., Chabot J., Nowyrod R., Greenberg S., Kuang W.J.,  
RA Leung D.W., Hayward J.R., Kistel W., Heath M., Brett J., Stern D.M.,  
RT "Characterization of a novel tumor-derived cytokine. Endothelial-  
RT monocyte activating polypeptide II.",  
RL J. Biol. Chem. 269:25106-25119 (1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Pancreas;  
RX MEDLINE=22388257; Straubeberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heish F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stepien M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Uedin T.B., Toshimiyuki S., Carninci P., Prange C.,  
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bock S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,  
RA Villalobos D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Faley J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whitting M., Madan A., Young A.C., Snevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield J.S.N., Krzywinski M.I., Skalska U., Small D.E.,  
RA Schenck A., Schein J.E., Jones S.J.M., Marra M.A.,  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences.",  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
RN [1-1] SMILLARITY: Contains 1 tRNA-binding domain.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and that statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>)

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
CC -----  
DR EMBL: U0117; AAA62202.1; -.  
DR EMBL: BC014051; AAH14051.1; -.  
DR PDB; 1E72; 06-FEB-01.  
DR PDB; 1E07; 06-SEP-00.  
DR PDB; 1FLO; 07-FEB-01.  
DR Genew; HGNC:10648; SCYE1.  
DR MIM; 603605; -.  
DR GO; GO:0005615; C:extracellular space; TAS.  
DR GO; GO:0005125; F:cytokine activity; TAS.  
DR GO; GO:0000049; F:tRNA binding activity; TAS.  
DR GO; GO:0006418; P:amino acid activation; TAS.  
DR GO; GO:0006935; P:chemotaxis; TAS.  
DR GO; GO:0006954; P:inflammatory response; TAS.  
DR InterPro; IPR002547; tRNA\_bind.  
DR Pfam; PF01588; tRNA\_bind\_1.  
DR PIRSF; PIRSF005381; EMAP11; 1.  
DR PROSITE; PS50886; TRBD; 1.  
KM Protein biosynthesis; RNA-binding; tRNA-binding; Cytokine;  
KM 3D-structure.  
FT PROPR 1 146  
FT CHAIN 147 312  
FT DOMAIN 151 252  
FT CONFLICT 79 79 A -> P (IN REF. 2).  
SQ SEQUENCE 312 AA; 34326 MW; 946310A0216F7587 CRC64;  
Query Match 100.0%; Score 516; DB 1; Length 312;  
Best Local Similarity 100.0%; Pred. No. 9e-31;  
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MANNDAVLRKLEKGAADQIIIEYLKQOVSILKEKALIQATLRREKKLRVENAKLKEIE 60  
DB 1 MANNDAVLRKLEKGAADQIIIEYLKQOVSILKEKALIQATLRREKKLRVENAKLKEIE 60  
QY 61 ELKQELIQAEIQNGVQKQIAFP--SGTPHANSVSENVYQSTAVTVTVSSG 108  
DB 61 ELKQELIQAEIQNGVQKQIAFP--SGTPHANSVSENVYQSTAVTVTVSSG 108  
RESULT 2  
MCAL CRIGR STANDARD; PRT; 359 AA.  
AC 054873;  
DT 16-OCT-2001 (Rel. 40; Last sequence update)  
DT 16-OCT-2001 (Rel. 40; Last sequence update)  
DT 28-FEB-2003 (Rel. 41; Last annotation update)  
DE Multisynthetase complex auxiliary component p43 [Contains:  
DE Endothelial-monocyte activating polypeptide II (EMAP-II) (Small  
DE Inducible cytokine subfamily E member 1)].  
GN SCYE1.  
OS Cricetus griseus (Chinese hamster).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
OC Cricetus.  
OX NCBI\_TaxID=10029;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE=98070438; Pubmed=9405472;  
RA Quevillon S., Agou F., Robinson J.-C., Mirande M.;  
RT "The p43 component of the mammalian multi-synthetase complex is likely  
RT to be the precursor of the endothelial monocyte-activating polypeptide  
RT II cytokine".  
RT J. Biol. Chem. 272:32573-32579 (1997).  
CC -1- SUBUNIT: COMPONENT OF THE MULTISYNTHETASE COMPLEX WHICH IS  
CC COMPOSED OF A BIFUNCTIONAL GLUTAMYL-PROLYL-TRNA SYNTHETASE, THE  
CC MONOSPECIFIC ISOLUCYL, LEUCYL, GLUTAMINYL, METHIONYL, LYSYL,  
CC ARGINYL, AND ASPARTYL-TRNA SYNTHETASES AS WELL AS THREE AUXILIARY  
CC PROTEINS, P18, P48 AND P43.  
CC -1- SIMILARITY: Contains 1 tRNA-binding domain.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: AF021800; AAB95207.1; -.  
DR InterPro; IPR002547; tRNA\_bind.  
DR Pfam; PF01588; tRNA\_bind\_1.  
DR PIRSF; PIRSF005381; EMAP11; 1.  
DR PROSITE; PS50886; TRBD; 1.  
KM Protein biosynthesis; RNA-binding; tRNA-binding; Cytokine.  
KM DOMAIN 138 299  
SQ SEQUENCE 359 AA; 39601 MW; 4B86BD1B65D72C23 CRC64;  
Query Match 82.4%; Score 425; DB 1; Length 359;  
Best Local Similarity 84.3%; Pred. No. 3.9e-24;  
Matches 91; Conservative 8; Mismatches 7; Indels 2; Gaps 1;  
QY 1 MANNDAVLRKLEKGAADQIIIEYLKQOVSILKEKALIQATLRREKKLRVENAKLKEIE 60  
DB 47 MANNDAVLRKLEKGAADQIIIEYLKQOVSILKEKALIQATLRREKKLRVENAKLKEIE 106  
QY 61 ELKQELIQAEIQNGVQKQIAFP--SGTPHANSVSENVYQSTAVTVTVSSG 106  
DB 107 ELKQELIQAEIQNGVQKQIAFPVQSDTPVQASAVSTVSIVQSTVSSTIS 154  
RESULT 3  
MCAL MOUSE STANDARD; PRT; 310 AA.  
AC P31230; Q60659;  
DT 01-JUL-1993 (Rel. 26; Last sequence update)  
DT 01-JUL-1993 (Rel. 26; Last sequence update)  
DT 15-SEP-2003 (Rel. 42; Last annotation update)  
DE Multisynthetase complex auxiliary component p43 [Contains:  
DE Endothelial-monocyte activating polypeptide II (EMAP-II) (Small  
DE Inducible cytokine subfamily E member 1)].  
GN SCYE1 OR EMAP2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE=95014290; Pubmed=7929199;  
RA Kao J., Houck K., Fan Y., Haehnel I., Libutti S.K., Kayton M.L.,  
RA Grikscheit T., Chabot J., Nowyrod R., Greenberg S., Kuang W.J.,  
RA Leung D.W., Hayward J.R., Kissel W., Heath M., Brett J., Stern D.M.;  
RT "Characterization of a novel tumor-derived cytokine. Endothelial-  
RT monocyte activating polypeptide II".  
RT J. Biol. Chem. 269:25106-25119 (1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22388257; Pubmed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Cantucci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., Weisman P.J., McKernan K.D., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hultky S.W.,  
RA Villalon D.K., Murry D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Buterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;



RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [3]  
RP SEQUENCE OF 145-164 FROM N.A.  
RX MEDLINE=93015897; PubMed=1400342;  
RA Kao J., Ryan U., Bretz G., Chen Y., Shen H., Fan Y.-G., Godman G.,  
FAmilletti P.C., Wang F., Pan Y.-C.E., Stern D., Claus M.;  
RT "Endothelial monocyte-activating polypeptide II. A novel  
RT tumor-derived polypeptide that activates host-response mechanisms.";  
RL J. Biol. Chem. 267:20239-20247(1992).  
RN [4]  
RP CHARACTERIZATION.  
RX MEDLINE=94193665; PubMed=7545917;  
RA Kao J., Fan Y., Haehnel I., Bretz J., Greenberg S., Claus M.,  
RA Kayton M., Houck K., Kiesel W., Seifried R., Burnier J., Stern D.;  
RT "A peptide derived from the amino terminus of endothelial-monocyte-  
RT activating polypeptide II modulates mononuclear and polymorphonuclear  
RT leukocyte functions, defines an apparently novel cellular interaction  
RT site, and induces an acute inflammatory response.";  
RL J. Biol. Chem. 269:9774-9782(1994).  
CC -1- FUNCTION: ALTERS ENDOTHELIAL AND MONOCYTE FUNCTIONS, INDUCES THE  
CC MIGRATION OF MONOCYTES AND GRANULOCYTES, AND INDUCES AN  
CC INFLAMMATORY RESPONSE IN THE MOUSE FOOTPAD MODEL. EMAP II ELICITS  
CC A PHLOGOGENIC RESPONSE AND, POTENTIALLY, AUGMENTS THE EFFECTS OF  
CC THE OTHER TUMOR-DERIVED CYTOKINES.  
CC -1- SUBUNIT: Monomer.  
CC -1- SIMILARITY: Contains 1 trna-binding domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U10118; AAG62203.1; -;  
DR EMBL; BC002054; AA02054.1; -;  
DR PIR; A55053; A55053.  
DR MGI; 102774; Scyl1.  
DR InterPro; IPR002547; tRNA\_bind.  
DR Pfam; PF01588; tRNA\_bind.1.  
DR PIRSF; PIRSF005381; EMBP1.1.  
DR PROSITE; PS50886; TRBD.1.  
DR Protein biosynthesis; RNA-binding; tRNA-binding; Cytokine.  
FT PROPEP 1 144  
FT CHAIN 145 310  
FT DOMAIN 149 250  
FT SEQUENCE 310 AA; 33997 MW; A2F8FF52A33D03A0 CRC64;  
Query Match 76.6%; Score 395; DB 1; Length 310;  
Best Local Similarity 77.6%; Pred. No. 5e-22;  
Matches 83; Conservative 8; Mismatches 16; Indels 0; Gaps 0;  
QY 1 MANDAVYKRLKQGAADQIIIEYKQOVSLKEKALIQATLRREKLRVNAKIKKEIE 60  
DB 1 MANDAVYKRLKQGAADQIIIEYKQOVSLKEKALIQATLRREKLRVNAKIKKEIE 60  
QY 61 ELKQELIAEIHNGEYVRVLSPLQNTASBESVVSPTAVTAS 107  
DB 61 ELKQELIAEIHNGEYVRVLSPLQNTASBESVVSPTAVTAS 107  
RESULT 4  
MCPB\_BACSU STANDARD; PRT; 662 AA.  
AC P39215;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DR Methyl-accepting chemotaxis protein mcpb (H3).

GN MCPB.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxId=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=168 / 011085;  
RA Azevedo V., Bouteiro M.G., Bessieres P., Bolotin A., Borchert S.,  
RA Borries R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,  
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,  
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
RA Enlian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
RA Fritz C., Fujita M., Fujita Y., Fuma S., Gallizi A., Galleron N.,  
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,  
RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,  
RA Hilbert H., Holstappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,  
RA Joris B., Karamata D., Kasahara Y., Klarr-Binchard M., Klein C.,  
RA Kodayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
RA Kunita K., Lapius A., Lardinois S., Lauber J., Lazarevic V.,  
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigne C.,  
RA Medina N., Melillo R.P., Mizuno M., Mostl D., Nakai S., Noback M.,  
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudga B., Park S.H.,  
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
RA Pressac E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadle Y.,  
RA Sager T., Scallan E., Schleich S., Schroeter R., Scoffone F.,  
RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Solido B.,  
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
RA Takeuchi M., Takakoshi A., Tanaka T., Terpetra P., Tognoni A.,  
RA Tosato V., Uchiyama S., Vandenbol M., Vanher F., Vassarotti A.,  
RA Viari A., Wambuit R., Wedler E., Wedler H., Welterneger T.,  
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
RA Yoshida K., Yoshikawa H.F., Zumestein E., Yoshikawa H., Danchin A.;  
RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
RT subtilis.";  
RL Nature 390:249-256(1997).  
CC -1- FUNCTION: RESPONSIBLE FOR DETECTING A SUBSET OF AMINO ACIDS THAT  
CC INCLUDES ASPARAGINE, ASPARTATE, GLUTAMINE, AND HISTIDINE.  
CC -1- FUNCTION: CHEMOTACTIC-SIGNAL TRANSDUCERS RESPOND TO CHANGES IN THE  
CC CONCENTRATION OF ATTRACTANTS AND REPELLENTS IN THE ENVIRONMENT, AND  
CC TRANSDUCE A SIGNAL FROM THE OUTSIDE TO THE INSIDE OF THE CELL, AND  
CC FACILITATE SENSORY ADAPTATION THROUGH THE VARIATION OF THE LEVEL  
CC OF METHYLATION. ALL AMINO ACIDS SERVE AS ATTRACTANTS IN  
CC B. SUBTILIS, THEY APPEAR TO CAUSE AN INCREASE IN THE TURNOVER  
CC METHYL GROUPS, LEADING TO METHYLATION OF AN UNIDENTIFIED ACCEPTOR,  
CC WHILE REPELLENTS HAVE BEEN SHOWN TO CAUSE A DECREASE IN METHYL  
CC GROUP TURNOVER. THE METHYL GROUPS ARE ADDED BY A METHYLTRANSFERASE  
CC AND REMOVED BY A METHYLERASE.  
CC -1- SIMILARITY: Contains 1 methyl-accepting transducer domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; L29189; AAA20554.1; -;  
DR EMBL; Z99119; CAB15104.1; -;  
DR EMBL; Z99120; CAB15115.1; -;

DR PIR; A54078; A54078.  
 DR HSP; P02942; 1Q07.  
 DR Subtilisin; BG10859; mcpB.  
 DR InterPro; IPR004010; Cache.  
 DR InterPro; IPR004089; Chmtaxis\_transd.  
 DR InterPro; IPR003660; HAMP.  
 DR InterPro; IPR003122; TazH.  
 DR Pfam; PF02743; Cache; 1.  
 DR Pfam; PF00672; HAMP; 1.  
 DR Pfam; PF00015; MCPsignal; 1.  
 DR Pfam; PF02203; TazH; 1.  
 DR SMART; SM00304; HAMP; 1.  
 DR SMART; SM00283; MA; 1.  
 DR SMART; SM00319; TazH; 1.  
 DR PROSITE; PSS0111; CHEMOTAXIS\_TRANSDUC\_2; 1.  
 DR PROSITE; PSS0885; HAMP; 1.  
 DR Chemotaxis; Transducer; Transmembrane; Methylation; Complete proteome.  
 DR DOMAIN; 1; 16; CYTOPLASMIC (POTENTIAL).  
 DR TRANSMEM; 17; 37; POTENTIAL.  
 DR DOMAIN; 38; 282; EXTRACELLULAR (POTENTIAL).  
 DR TRANSMEM; 283; 303; POTENTIAL.  
 DR DOMAIN; 304; 662; CYTOPLASMIC (POTENTIAL).  
 DR DOMAIN; 375; 611; HAMP.  
 DR MOD\_RES; 371; 371; METHYL-ACCEPTING TRANSDUCER.  
 DEAMIDATION AND METHYLATION (BY SIMILARITY).  
 METHYLATION (BY SIMILARITY).  
 FT MOD\_RES; 630; 630; METHYLATION (BY SIMILARITY).  
 FT MOD\_RES; 637; 637; METHYLATION (BY SIMILARITY).  
 SQ SEQUENCE; 662 AA; 71885 MW; 98DD181PB46900BC CRC64;

Query Match 18.1%; Score 93.5; DB 1; Length 662;  
 Best Local Similarity 29.8%; Pred. No. 6.6; Indels 27; Gaps 4;  
 Matches 34; Conservative 16; Mismatches 37;

QY 5 DAVLKLEQKGAADQIIEYK--QGVSLKEKALIQ-----TLREKQLR 49  
 DB 464 EAVVKGLEKSKQITSLRVINGIADQTNLALNAEARGSGRGFSVAEVRKLA 523  
 QY 50 VERNAKKEIEBELKQELIQAEIONGVQKQAPPSGTPPHANSMVSENYGSTAVT 103  
 DB 524 VQSDAKSEIKELIKQELIV-AEID-----TSLHKEVQVQVQSGIYVVT 565

RESULT 5  
 YL17\_CABEL STANDARD; PRT; 1130 AA.

AC Q1102;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical 131.5 kDa protein CO2F12.7 in chromosome X.  
 GN CO2F12.7.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 NCBI\_Taxid=6239;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Miller N.;  
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: WEAK, TO MYOSINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; U41545; AAK39135.1; -.  
 DR PIR; T34081; T34081.  
 DR WormPep; CO2F12.7; CE03901.  
 DR Hypothetical protein; Coiled coil.  
 FT DOMAIN; 121; 779; COILED COIL (POTENTIAL).  
 FT DOMAIN; 805; 1061; COILED COIL (POTENTIAL).  
 SQ SEQUENCE; 1130 AA; 131485 MW; B0FD2EF3D99FB09 CRC64;

Query Match 18.0%; Score 93; DB 1; Length 1130;  
 Best Local Similarity 24.6%; Pred. No. 12;  
 Matches 29; Conservative 30; Mismatches 33; Indels 26; Gaps 4;

QY 2 ANDAVLKLEQKGAADQIIEYKQVSLKEKALIQTLREK-----KLRYENAKL 55  
 DB 691 SKXNDLLEELSK-----KNRIEHLKQELIKQELIKSTKETKQSELEKTLAQLEIDNSK 746  
 QY 56 KKEIEEL-----KQELIQAEIONGVQKQAPPSGTPPHANSMVSENYGSTAVT 99  
 DB 747 SQIEKRLHVRVNDMDQGTIDELVKNE--IKTISAKTQALLSENTVESITYLAS 802

RESULT 6  
 SMC\_METUA STANDARD; PRT; 1169 AA.

AC Q59037;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Chromosome partition protein smc homolog.  
 GN M01643.  
 OS Methanococcus jannaschii.  
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
 OC Methanocaldococcaceae; Methanocaldococcus.  
 NCBI\_Taxid=2190;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;  
 RX MEDLINE=96337999; PubMed=8688087;  
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,  
 RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,  
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,  
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,  
 RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus  
 jannaschii.";  
 RL Science 273:1058-1073 (1996).

RN [2]  
 RP REVISIONS.  
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,  
 RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,  
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,  
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,  
 RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN CHROMOSOME STRUCTURE AND  
 CC PARTITIONING. ESSENTIAL FOR CHROMOSOME PARTITION (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE SMC FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR TIGR; M1643; --  
DR InterPro; IPR003439; ABC transporter.  
DR InterPro; IPR005289; GTP-binding\_dom.  
DR InterPro; IPR003405; SMC\_C.  
DR InterPro; IPR003395; SMC\_N.  
DR Pfam; PF02483; SMC\_C.1.  
DR Pfam; PF02463; SMC\_N.1.  
DR TIGR/Pfam; TIGR00650; MG42; 1.  
KW Hypothetical protein; ATP-binding; Coiled coil; Complete proteome.  
FT NP BIND 31 38 ATP (POTENTIAL).  
FT DOMAIN 160 521 COILED COIL (POTENTIAL).  
FT DOMAIN 673 1032 COILED COIL (POTENTIAL).  
SQ SEQUENCE 1169 AA; 136634 MW; B63CE34E4C03F36 CRC64;  
Query Match 17.8%; Score 92; DB 1; Length 1169;  
Best Local Similarity 26.6%; Pred. No. 15;  
Matches 25; Conservative 24; Mismatches 25; Indels 20; Gaps 4;  
OY 3 NNDVAVLEKLEQKAGADQIIIEYLKQ---QVSLKEKALQATYREKKLRVENAKLKE 58  
DB 281 NN--IINELNKGNE--EVLHLKSKIKELVEIENDKVLDSINELKKVEIENKKKE 336  
OY 59 IEELEKQELIQ-----AEIQNGVKQIAP 80  
DB 337 IKETQKLIENRDSIIKEEQQIKELIEKIKMLNY 370  
RESULT 7  
MARG\_HUMAN STANDARD; PRT; 162 AA.  
ID MARG\_HUMAN  
AC O15525;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Transcription factor MafG (V-maf musculoaponeurotic fibrosarcoma oncogene homolog G) (hMaf).  
GN MAFG.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
NCBI\_TaxID=9606;  
RX SEQUENCE FROM N.A.  
RX MEDLINE=97341189; PubMed=9195958;  
RA Martini M.G., Chan K., Casula L., Kan Y.W., Cao A., Moi P.;  
RT "hMaf, a small human transcription factor that heterodimerizes specifically with Nrf1 and Nrf2."  
RL J. Biol. Chem. 272:16490-16497(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RP TISSUE=Muscle;  
RX MEDLINE=97309414; PubMed=9166829;  
RA Blank V., Kim M.J., Andrews N.C.;  
RT "Human MafG is a functional partner for p45 NF-E2 in activating globin gene expression."  
RL Blood 89:3925-3935(1997).  
RN [3]  
RP SEQUENCE FROM N.A.  
RP TISSUE=Muscle;  
RX MEDLINE=97432833; PubMed=9286713;  
RA Blank V., Knoll J.H.M., Andrews N.C.;  
RT "Molecular characterization and localization of the human MafG gene."  
RL Genomics 44:147-149(1997).  
RN [4]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97294601; PubMed=9150357;  
RA Toki T., Itoh J., Kitazawa J., Arai K., Hatakeyama K., Akasaka J., Igarashi K., Nomura N., Yokoyama M., Yamamoto M., Ito E.;  
RT "Human small Maf proteins form heterodimers with CNC family transcription factors and recognize the NF-E2 motif."  
RL Oncogene 14:1901-1910(1997).  
RN [5]

RP SEQUENCE FROM N.A.  
RA Ito E., Toki T.;  
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.  
RN [6]  
RP SEQUENCE FROM N.A.  
RP TISSUE=Lung;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Straubeberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenman C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loguettano N.A., Peters G.J., Abramson R.D., Mullaly S.J., Bosak S.A., McMan P.J., McKernan K.J., Malek J.A., Gunnaracne P.H., Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Huiyk S.W., Villalón D.K., Wuzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Faley J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences."  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
CC -1- FUNCTION: SINCE THEY LACK A PUTATIVE TRANSCRIPTION DOMAIN, THE SMALL MAFs BEHAVE AS TRANSCRIPTIONAL REPRESSORS WHEN THEY DIMERIZE AMONG THEMSELVES. HOWEVER, THEY SEEM TO SERVE AS TRANSCRIPTIONAL ACTIVATORS BY DIMERIZING WITH OTHER (USUALLY LARGER) BASIC-ZIPPER PROTEINS AND RECRUITING THEM TO SPECIFIC DNA-BINDING SITES. SMALL MAF PROTEINS HETERODIMERIZE WITH POS AND MAY ACT AS COMPETITIVE REPRESSORS OF THE NF-E2 TRANSCRIPTION FACTOR. TRANSCRIPTION FACTOR, COMPONENT OF ERYTHROID-SPECIFIC TRANSCRIPTION FACTOR NF-E2. ACTIVATES GLOBIN GENE EXPRESSION WHEN ASSOCIATED WITH NF-E2.  
CC -1- SUBUNIT: HOMO- OR HETERODIMER. IT CAN ALSO FORM HIGH AFFINITY HETERODIMERS WITH MEMBERS OF THE NFE2/CNC-BZIP FAMILY: NRE1, NRE2 AND P45-NRE2.  
CC -1- SUBCELLULAR LOCATION: Nuclear.  
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN SKELETAL MUSCLE. ALSO EXPRESSED IN HEART AND BRAIN.  
CC -1- SIMILARITY: Belongs to the bZIP family. Maf subfamily.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.ebi.ac.uk/announcements> or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
CC -----  
DR EMBL; Y11514; CAAT2284.1; --  
DR EMBL; U84249; AAC51737.1; --  
DR EMBL; AF059195; AAC14427.1; --  
DR EMBL; BC012327; AAH12327.1; --  
DR TRANSFAC; T04870; --  
DR Genew; HGNC:6781; MAFG.  
DR MIM; 602020; --  
DR InterPro; IPR004827; TF bZIP.  
DR InterPro; IPR004826; TF Maf.  
DR Pfam; PF03131; bZIP\_Maf\_1.  
DR SMART; SM00358; BRIZ; 1.  
DR PROSITE; PS50217; bZIP\_1.  
DR PROSITE; PS00036; bZIP\_BASIC; FALSE NEG.  
KW Transcription regulation; DNA-binding; Repressor; Nuclear protein.  
FT DNA BIND 53 83 BASIC MOTIF.  
FT DOMAIN 86 114 LEUCINE-ZIPPER.  
SQ SEQUENCE 162 AA; 17849 MW; E49F1FBA230F8D30 CRC64;  
Query Match 17.7%; Score 91.5; DB 1; Length 162;  
Best Local Similarity 30.1%; Pred. No. 2.2;  
Matches 31; Conservative 21; Mismatches 44; Indels 7; Gaps 3;

QY 9 KRLEKGAADQIIEYLKQVSLKEKAILQATLREKKLRVENAKLKEIEEL--KOEL 66  
 DB 57 RTLKNRGVYASCRVKTQKELEKQKELQ---QEVKGLASNSMKLELDALRSKYEA 113  
 QY 67 IQAEIONGVKQIAFPSCGTPPLA--NSMSENVIQSTRAVTVSS 107  
 DB 114 LQTFARTVARSVPVAPARGPLAAGLPLVPGKVAATSVITIVKS 156

## RESULT 8

PS4\_ENTFC STANDARD; PRT; 516 AA.  
 ID PS4\_ENTFC STANDARD; PRT; 516 AA.  
 AC P13692;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-NOV-1991 (Rel. 20, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE P54 protein precursor.  
 OS Enterococcus faecium (Streptococcus faecium).  
 OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.  
 NCBI\_TaxId=1352;  
 RX MEDLINE=69385998; PubMed=2780297;  
 RA Feuer P., Moesch H.-U., Solioz M.;  
 RT "A protein of unusual composition from Enterococcus faecium";  
 RL Nucleic Acids Res. 17:6724-6724(1989).  
 CC -1- SUBCELLULAR LOCATION: Attached to the cell wall.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C40.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; X16421; CAA34442.1; ALT\_INIT.  
 DR PIR; S05542; S05542.  
 DR MEROPS; C40; UPW; -  
 DR InterPro; IPR000064; NLP\_C\_P60.  
 DR Pfam; PF00877; NLP\_C\_P60; 1.  
 KW Signal; Cell wall.  
 FT SIGNAL 1 27 POTENTIAL.  
 FT CHAIN 28 516 P54 PROTEIN.  
 FT ACT\_SITE 429 429 POTENTIAL.  
 SQ SEQUENCE 516 AA; 54596 MW; 402ECAA439846D26 CRC64;  
 Query Match 17.7%; Score 91.5; DB 1; Length 516;  
 Best Local Similarity 30.3%; Pred. No. 7.2;  
 Matches 33; Conservative 28; Mismatches 37; Indels 11; Gaps 4;  
 QY 5 DAVLKLEKGAADQIIEYLKQVSLKEKAILQATLREKKLRVENAKLKEIEELKQ 64  
 DB 38 DKRIADLQNGQASQISQIEALKEGVSAINTKA--QDLTKQPTLRKESADLQKEIKDQ 95  
 QY 65 --ELIQAELIONGVKQIAFPSCGTP-----LHANSMSSENVISQSTRAVTV 105  
 DB 96 RLEKREKATIQKQARETOVKNTSSNYIDAVLADSL-ADAVGRIQAMSTI 143

## RESULT 9

MAFG\_CHICK STANDARD; PRT; 162 AA.  
 ID MAFG\_CHICK STANDARD; PRT; 162 AA.  
 AC Q90889;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Transcription factor Mafg (V-maf musculoaponeurotic fibrosarcoma oncogene homolog G).  
 OS Gallus gallus (Chicken).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxId=9031;  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=95198738; PubMed=7891713;  
 RX Kataoka K., Igatahshi K., Itoh K., Fujiwara K.T., Noda M.,  
 RA Yamamoto M., Nishizawa M.;  
 RT "Small Maf proteins heterodimerize with Fos and may act as competitive repressors of the NF-E2 transcription factor";  
 RL Mol. Cell. Biol. 15:2180-2190(1995).  
 CC -1- FUNCTION: SINCE THEY LACK A PUTATIVE TRANSDOMAIN, THE SMALL MAFs BEHAVE AS TRANSCRIPTIONAL REPRESSORS WHEN THEY DIMERIZE AMONG THEMSELVES. HOWEVER, THEY SEEM TO SERVE AS TRANSCRIPTIONAL ACTIVATORS BY DIMERIZING WITH OTHER (USUALLY LARGER) BASIC-ZIPPER PROTEINS AND RECRUITING THEM TO SPECIFIC DNA-BINDING SITES. SMALL MAF PROTEINS HETERODIMERIZE WITH FOS AND MAY ACT AS COMPETITIVE REPRESSORS OF THE NF-E2 TRANSCRIPTION FACTOR. TRANSCRIPTION FACTOR, COMPONENT OF ERYTHROID-SPECIFIC TRANSCRIPTION FACTOR NF-E2.

CC -1- SUBUNIT: HOMO- OR HETERODIMER.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- SIMILARITY: Belongs to the bZIP family. Maf subfamily.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; D28602; BA05939.1; -  
 DR PIR; I50378; A56254.  
 DR TRANSFAC; T01437; -  
 DR InterPro; IPR004827; TF\_bZIP.  
 DR Pfam; PF03131; bZIP\_Maf.1.  
 DR SMART; SM00338; BRLZ; 1.  
 DR PROSITE; PS0217; BZIP; 1.  
 DR PROSITE; PS00036; BZIP\_BASIC; FALSE NEG.  
 KW Transcription regulation; DNA-binding; Repressor; Nuclear protein.  
 FT DNA\_BIND 53 83 BASIC MOTIF.  
 FT DOMAIN 86 114 LEUCINE-ZIPPER.  
 SQ SEQUENCE 162 AA; 18077 MW; C7B0FCD18800696C CRC64;

Query Match 17.5%; Score 90.5; DB 1; Length 162;  
 Best Local Similarity 27.6%; Pred. No. 2.7;  
 Matches 29; Conservative 23; Mismatches 42; Indels 11; Gaps 3;  
 QY 9 KRLEKGAADQIIEYLKQVSLKEKAILQATLREKKLRVENAKLKEIEELKQELQ 68  
 DB 57 RTLKNRGVYASCRVKTQKELEKQKELQ---QEVKGLASNSMKLELDALRSK--Y 111  
 QY 69 AEIONGVKQIAFPSCGTPPLA-----NSMSENVIQSTRAVTVSS 107  
 DB 112 EALQNFARTVARSVPVAPARGPLTSMGRLVPGKVAATSVITIVKS 156

## RESULT 10

KFSC\_MOUSE STANDARD; PRT; 956 AA.  
 ID KFSC\_MOUSE STANDARD; PRT; 956 AA.  
 AC P28738; Q922F8;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Kinesin heavy chain isoform 5C (Kinesin heavy chain neuron-specific 2).  
 GN KFSC OR NKG2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

OX NCB1_TaxID=10090;
RN [1]
RP PRELIMINARY SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Brain;
RA Kato K.;
RT "A collection of cDNA clones with specific expression patterns in
   mouse brain."
RL Eur. J. Neurosci. 2:704-711(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=9900842; PubMed=9782088;
RA Xia C., Rahman A., Yang Z., Goldstein L.S.B.;
RT "Chromosomal localization reveals three kinesin heavy chain genes in
   mouse."
RL Genomics 52:209-213(1998).
CC -1- FUNCTION: KINESIN IS A MICROTUBULE-ASSOCIATED FORCE-PRODUCING
CC PROTEIN THAT MAY PLAY A ROLE IN ORGANELLE TRANSPORT.
CC -1- SUBUNIT: OLIGOMER COMPOSED OF TWO HEAVY CHAINS AND TWO LIGHT
CC CHAINS.
CC -1- DOMAIN: COMPOSED OF THREE STRUCTURAL DOMAINS: A LARGE GLOBULAR N-
CC TERMINAL DOMAIN WHICH IS RESPONSIBLE FOR THE MOTOR ACTIVITY OF
CC KINESIN (IT HYDROLYZES ATP AND BINDS MICROTUBULE), A CENTRAL
CC ALPHA-HELICAL COILED COIL DOMAIN THAT MEDIATES THE HEAVY CHAIN
CC DIMERIZATION, AND A SMALL GLOBULAR C-TERMINAL DOMAIN WHICH
CC INTERACTS WITH OTHER PROTEINS (SUCH AS THE KINESIN LIGHT CHAINS),
CC VESICLES AND MEMBRANOUS ORGANELLES.
CC -1- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN
CC SUBFAMILY.
CC -1- CAUTION: REF.1 SEQUENCE SEEMS TO HAVE BEEN A HYBRID OF A KIF5A
CC AND A KIF5C SEQUENCE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; X61435; CAA43677.1; -.
DR EMBL; AF067180; AAC79804.1; -.
DR HSSP; P56536; 2KIN.
DR MGD; MGI:1098269; Klf5c.
DR GO; GO:0005515; F:protein binding activity; IPI.
DR GO; GO:0008045; P:motor axon guidance; IMP.
DR InterPro; IPR001752; kinesin_motor.
DR Pfam; PF00225; kinesin; 1.
DR PRINTS; PR00380; KINESINHEAVY.
DR SMART; SM00129; KISC; 1.
DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
DR Motor protein; Microtubules; ATP-binding; Coiled coil.
FT DOMAIN 1 386 KINESIN_MOTOR (BY SIMILARITY).
FT DOMAIN 406 923 COILED COIL.
FT DOMAIN 859 956 GLOBULAR.
FT DOMAIN 174 315 MICROTUBULE-BINDING.
FT NP_BIND 86 93 ATP (BY SIMILARITY).
SO SEQUENCE 956 AA; 109240 MW; D5A8C701A2911AB2 CRC64;
Query Match 17.4%; Score 90; DB 1; Length 956;
Best Local Similarity 26.1%; Pred. No. 17;
Matches 29; Conservative 25; Mismatches 43; Indels 14; Gaps 4;
OY 6 AVLKRLBQKGAAD---QIIEYLKQVSLLEKAILQATLEBEK-----LRYENAKL 55
Db 422 SLVRLDLDKDEINQSOQLAEKLKQ--MLDQDELLASTRDYKIQEQLFRLLQIENBAA 479
OY 56 KKEIEELNGEILQAEIQNGVQKQIAPSPGPHANSMVENYQSAVTVTS 106
Db 480 KDEVEVQLQAELEAVNVDQKQSEVEDKT--RANQLTDELAKTKTTLLTT 528

```

```

RESULT 11
KIF5C HUMAN
ID_KIF5C HUMAN STANDARD; PRT; 957 AA.
AC 0602B2; 095079;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Kinesin heavy chain isoform 5C (Kinesin heavy chain neuron-specific
   2).
DN KIF5C OR NKHC2 OR KIAA0531.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCB1_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98290545; PubMed=9628581;
RA Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,
RA Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. IX.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro."
RL DNA Res. 5:31-39(1998).
RN [2]
RP SEQUENCE OF 355-585 FROM N.A.
RA Engelender S., Sharp A.H., Colomer V., Tokito M.K., Lanahan A.,
RA Wolley P., Holzbaur E.L.F., Ross C.A.;
RT "Huntingtin associated protein 1 (HAP1) interacts with the p150glued
RT subunit of dynein."
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: KINESIN IS A MICROTUBULE-ASSOCIATED FORCE-PRODUCING
CC PROTEIN THAT MAY PLAY A ROLE IN ORGANELLE TRANSPORT.
CC -1- SUBUNIT: OLIGOMER COMPOSED OF TWO HEAVY CHAINS AND TWO LIGHT
CC CHAINS.
CC -1- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN BRAIN, PROSTATE, AND
CC TESTIS, AND MODERATE EXPRESSION IN KIDNEY, SMALL INTESTINE, AND
CC OVARY.
CC -1- DOMAIN: COMPOSED OF THREE STRUCTURAL DOMAINS: A LARGE GLOBULAR N-
CC TERMINAL DOMAIN WHICH IS RESPONSIBLE FOR THE MOTOR ACTIVITY OF
CC KINESIN (IT HYDROLYZES ATP AND BINDS MICROTUBULE), A CENTRAL
CC ALPHA-HELICAL COILED COIL DOMAIN THAT MEDIATES THE HEAVY CHAIN
CC DIMERIZATION, AND A SMALL GLOBULAR C-TERMINAL DOMAIN WHICH
CC INTERACTS WITH OTHER PROTEINS (SUCH AS THE KINESIN LIGHT CHAINS),
CC VESICLES AND MEMBRANOUS ORGANELLES.
CC -1- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AB011103; BAA25457.1; -.
DR EMBL; AF010146; AAD01436.1; -.
DR HSSP; P56536; 2KIN.
DR Genew; HGNC:6325; KIF5C.
DR MIM; 604593; -.
DR GO; GO:0005871; C:kinesin complex; TAS.
DR GO; GO:0003777; F:microtubule motor activity; TAS.
DR GO; GO:0006996; P:organelle organization and biogenesis; TAS.
DR InterPro; IPR001752; kinesin_motor.
DR Pfam; PF00225; kinesin; 1.
DR PRINTS; PR00380; KINESINHEAVY.
DR SMART; SM00129; KISC; 1.
DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
DR Motor protein; Microtubules; ATP-binding; Coiled coil.
FT DOMAIN 1 386 KINESIN_MOTOR (BY SIMILARITY).
FT DOMAIN 406 923 COILED COIL.

```

FT DOMAIN 859 956 GLOBULAR.  
 FT DOMAIN 174 315 MICROVIBULE-BINDING.  
 FT NP BIND 86 93 ATP (BY SIMILARITY).  
 FT CONFLICT 355 360 TLNVI -> STNASH (IN REF. 2).  
 FT CONFLICT 583 585 EFT -> DRY (IN REF. 2).  
 SQ SEQUENCE 957 AA; 109494 MW; A9F25B1C994322A CRC64;  
 Query Match 17.4%; Score 90; DB 1; Length 957;  
 Best Local Similarity 26.1%; Pred. No. 17;  
 Matches 29; Conservative 25; Mismatches 43; Indels 14; Gaps 4;  
 QY 6 AVLKRLQKGAAD---QIIEYLKQVSLKEKAILQATREKK-----LRVNAKL 55  
 DB 423 STYRQDDKDDKDEINQOSLAELKQ--MDODELASTRDEYKIOEELRLQIENEA 480  
 QY 56 KKEIEELKQELIQAEIQNGVKQIAFPGGFLHANSWSENVISTAVTVSS 106  
 DB 481 KDEVKEVLQAELEAVNYDQKSGQVEDKT--RANEQLTDELAQKTTTLTT 529  
 RESULT 12  
 MARG\_MOUSE STANDARD; PRT; 162 AA.  
 ID MARG\_MOUSE  
 AC 054790;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Transcription factor Mafg (V-maf musculoaponeurotic fibrosarcoma  
 oncogene homolog G).  
 GN MAFG.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RP STRAIN=129/SvJ;  
 RX MEDLINE=98344009; PubMed=9679061;  
 RA Shavit J.A., Mochanashi H., Onodera K., Akasaka J., Yamamoto M.,  
 RA Engel J.D.;  
 RT "Impaired megakaryopoiesis and behavioral defects in mafg-null mutant  
 mice.";  
 RL Genes Dev. 12:2164-2174(1998).  
 RN [2]  
 RP STRUCTURE BY NMR OF 24-64.  
 RX MEDLINE=21912421; PubMed=11875518;  
 RA Kusunoki H., Mochanashi H., Katsunaka F., Mochanashi A., Yamamoto M.,  
 RA Tanaka T.;  
 RT "Solution structure of the DNA-binding domain of Mafg.";  
 RL Nat. Struct. Biol. 9:252-256(2002).  
 CC -1- FUNCTION: SINCE THEY LACK A PUTATIVE TRANSACTIVATION DOMAIN, THE  
 CC SMALL MAPS BEHAVE AS TRANSCRIPTIONAL REPRESSORS WHEN THEY DIMERIZE  
 CC AMONG THEMSELVES. HOWEVER, THEY SEEM TO SERVE AS TRANSCRIPTIONAL  
 CC ACTIVATORS BY DIMERIZING WITH OTHER (USUALLY LARGER) BASIC-ZIPPER  
 CC PROTEINS AND RECRUITING THEM TO SPECIFIC DNA-BINDING SITES. SMALL  
 CC MAP PROTEINS HETERODIMERIZE WITH FOS AND MAY ACT AS COMPETITIVE  
 CC REPRESSORS OF THE NF-E2 TRANSCRIPTION FACTOR. TRANSCRIPTION  
 CC FACTOR, COMPONENT OF ERYTHROID-SPECIFIC TRANSCRIPTION FACTOR NF-  
 CC E2. ACTIVATES GLOBIN GENE EXPRESSION WHEN ASSOCIATED WITH NF-E2  
 CC (BY SIMILARITY).  
 CC -1- SUBUNIT: HOMO-OR HETERODIMER (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).  
 CC -1- SIMILARITY: Belongs to the bzip family. Maf subfamily.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; AB009693; BAA24028.1; -.

DR PDB; 1KIV; 18-DEC-02.  
 DR MGD; MGJ:96911; Mafg.  
 DR GO; GO:0030534; P:regul. behavior; IMP.  
 DR GO; GO:0042127; P:regulation of cell proliferation; IMP.  
 DR InterPro; IPR004827; TF\_bzip.  
 DR InterPro; IPR004826; TF\_Maf.  
 DR Pfam; PF03131; bzip\_Maf; 1.  
 DR SMART; SM00338; BRLZ; 1.  
 DR PROSITE; PS00217; BZIP; 1.  
 DR PROSITE; PS00036; BZIP\_BASIC; FALSE NEG.  
 KW Transcription regulation; DNA-binding; Repressor; Nuclear protein;  
 KM 3D-structure.  
 FT DNA BIND 53 83 BASIC MOTIF.  
 FT DOMAIN 86 114 LEUCINE-ZIPPER.  
 SQ SEQUENCE 162 AA; 17876 MW; C7FF19614EB95C7D CRC64;  
 Query Match 17.3%; Score 89.5; DB 1; Length 162;  
 Best Local Similarity 30.1%; Pred. No. 3.1;  
 Matches 31; Conservative 21; Mismatches 44; Indels 7; Gaps 3;  
 QY 9 KRLQKGAADQIIEYLKQVSLKEKAILQATREKKLRVNAKLKKEIEEL--KQEL 66  
 DB 57 RLTKRGVYASGVKRVYQKELEKQKALQ---QEVETLASENSMKLELALSKYRA 113  
 QY 67 IQAEIQNGVKQIAFPGGFLHA--NSWSENVISTAVTVSS 107  
 DB 114 LQNFARTVARSVPAPRGFLAAGLPGKVAATSVITVKS 156  
 RESULT 13  
 KINH\_CAEEL STANDARD; PRT; 815 AA.  
 ID KINH\_CAEEL  
 AC P34540;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE kinesin heavy chain.  
 GN UNC-116 OR R05D3.7.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Pelodermidae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RP STRAIN=Bristol N2;  
 RX MEDLINE=94022338; PubMed=8105472;  
 RA Patel N., Thierry-Mieg D., Mancillas J.R.;  
 RT "Cloning by insertional mutagenesis of a cDNA encoding Caenorhabditis  
 RT elegans kinesin heavy chain.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 90:9181-9185(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Siddiqui S.S., Ali M.Y., Khan M.A.;  
 RL Submitted (Aug-1998) to the EMBL/Genbank/DBJ databases.  
 RN [3]  
 RC SEQUENCE FROM N.A.  
 RP STRAIN=Bristol N2;  
 RX MEDLINE=94150718; PubMed=7906398;  
 RA Wilson R., Ahnscough R., Anderson K., Baynes C., Berks M.,  
 RA Bonfield J., Burton J., Connell M., Copesey T., Cooper J., Coulson A.,  
 RA Craxton M., Dear S., Du Z., Durbin R., Favellio A., Fraser A.,  
 RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,  
 RA Johnston L., Jones M., Kershaw J., Kirsten J., Lalister N.,  
 RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,  
 RA Parsons J., Percy C., Rifken L., Koopa A., Saunders D., Shownkeen R.,  
 RA Sims M., Smalton N., Smith A., Smith M., Sonhammer E., Staden K.,  
 RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,  
 RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,  
 RA Woldman P.;  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 RT elegans.";  
 RL Nature 368:32-38(1994).  
 RN [4]

RP REVISIONS.  
RA Waterston R.;  
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: KINESIN IS A MICROTUBULE-ASSOCIATED FORCE-PRODUCING  
CC PROTEIN THAT MAY PLAY A ROLE IN ORGANELLE TRANSPORT.  
CC -1- SUBUNIT: OLIGOMER COMPOSED OF TWO HEAVY CHAINS AND TWO LIGHT  
CC CHAINS.  
CC -1- DOMAIN: COMPOSED OF THREE STRUCTURAL DOMAINS: A LARGE GLOBULAR N-  
CC TERMINAL DOMAIN WHICH IS RESPONSIBLE FOR THE MOTOR ACTIVITY OF  
CC KINESIN (IT HYDROLYSES ATP AND BINDS MICROTUBULE), A CENTRAL  
CC ALPHA-HELICAL COILED COIL DOMAIN THAT MEDIATES THE HEAVY CHAIN  
CC DIMERIZATION, AND A SMALL GLOBULAR C-TERMINAL DOMAIN WHICH  
CC INTERACTS WITH OTHER PROTEINS (SUCH AS THE KINESIN LIGHT CHAINS),  
CC VESICLES AND MEMBRANOUS ORGANELLES.  
CC -1- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN  
CC SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; L19120; AAA28155.1; -  
CC EMBL; AB017163; BAA32594.1; -  
CC EMBL; L07144; AAK21446.1; -  
CC PIR; S44868; S44868.  
CC HSSP; P33176; 1B62.  
CC WormPep; R05D3.7; CE25070.  
CC InterPro; IPR001752; kinesin\_motor.  
CC Pfam; PF00225; kinesin\_1.  
CC PRINTS; PR00380; KINESINHEAVY.  
CC SMART; SM00129; KISC; 1.  
CC DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.  
CC DR PROSITE; PS00067; KINESIN MOTOR DOMAIN2; 1.  
CC KW Motor protein; Microtubules; ATP-binding; Coiled coil.  
CC FT DOMAIN 1 259 KINESIN-MOTOR.  
CC FT DOMAIN 335 374 COILED COIL (POTENTIAL).  
CC FT DOMAIN 422 554 COILED COIL (POTENTIAL).  
CC FT NP BIND 695 785 COILED COIL (POTENTIAL).  
CC FT NP BIND 88 95 ATP (BY SIMILARITY).  
CC SQ SEQUENCE 815 AA; 91893 MW; 1B32718C3A7C254C CRC64;  
Query Match 17.3%; Score 89.5; DB 1; Length 815;  
Best Local Similarity 36.0%; Pred. No. 16;  
Matches 27; Conservative 22; Mismatches 17; Indels 9; Gaps 5;  
QY 9 KRLEQKGAEDQI--IEVLKQVSLTKKAILQATLR-EKKLIVENAKLKKEIEELKO 64  
DB 440 QQLDKKDEIKVSGSELEKTRQGV-LLOSEAL--GIMENBELIENNRFOKEAEDKQ 496  
QY 65 ELIQAEIONGVKQIA 79  
DB 497 E--GKEMMTALEEIA 509  
RESULT 14  
CLPB\_HELPY STANDARD; PRT; 856 AA.  
AC P71404;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE CLPB protein.  
GN CLPB OR HP0264.  
OS Helicobacter pylori (Campylobacter pylori).  
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;  
OC Helicobacteraceae; Helicobacter.  
OX NCBI\_TaxID=210;  
RN (1)  
RP SEQUENCE FROM N.A.

RC STRAIN=NCTC 11637;  
RA Allan E.;  
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=26695 / ATCC 700392;  
RX MEDLINE=97394467; PubMed=9252185;  
RA Tomb J.-F., White O., Kurland A.R., Clayton R.A., Sutton G.G.,  
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,  
RA Nelson K., Richardson J., Zhou L., Kirtress E.F., Peterson S.,  
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,  
RA McInerney J., Gocayne J.D., Lee N., Adams M.D., Hickey E.K.,  
RA Beyt D.E., Gocayne J.D., Utechtack T.R., Peterson J.D., Kelley J.M.,  
RA Cotton M.D., Feldman J.M., Fujii C., Bowman C., Wathey L., Wallin E.,  
RA Hooton W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,  
RA Venter J.C.;  
RT "The complete genome sequence of the gastric pathogen Helicobacter  
RT pylori.";  
RL Nature 388:539-547(1997).  
CC -1- SUBUNIT: THOUGHT TO BE AN ATPASE SUBUNIT OF AN INTRACELLULAR  
CC ATP-DEPENDENT PROTEASE (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE CLPB/CLPB FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; Y08238; CAA69406.1; -  
CC EMBL; AE000545; AAD07330.1; -  
CC PIR; H64552; H64552.  
CC TIGR; HP0264; -  
CC InterPro; IPR003593; AAA ATPase.  
CC InterPro; IPR003959; AAA ATPase center.  
CC InterPro; IPR001270; Chaprinin\_clpA/B.  
CC InterPro; IPR004176; Clp\_N.  
CC Pfam; PF00004; AAA; 2.  
CC Pfam; PF02861; Clp\_N; 2.  
CC PRINTS; PR00300; CLPPTOTASEA.  
CC DR SMART; SM00382; AAA; 2.  
CC DR PROSITE; PS00870; CLPB 1; 1.  
CC DR PROSITE; PS00871; CLPB 2; 1.  
CC KW Chaperone; ATP-binding; Repeat; Complete proteome.  
CC FT DOMAIN 159 406 1.  
CC FT NP BIND 531 722 11.  
CC FT NP BIND 204 211 ATP (POTENTIAL).  
CC FT NP BIND 605 612 ATP (POTENTIAL).  
CC FT CONFLICT 14 14 T -> A (IN REF. 1).  
CC FT CONFLICT 33 33 M -> L (IN REF. 1).  
CC FT CONFLICT 59 59 Q -> E (IN REF. 1).  
CC FT CONFLICT 62 62 R -> K (IN REF. 1).  
CC FT CONFLICT 79 79 S -> N (IN REF. 1).  
CC FT CONFLICT 103 103 R -> T (IN REF. 1).  
CC FT CONFLICT 119 119 G -> S (IN REF. 1).  
CC FT CONFLICT 131 131 A -> T (IN REF. 1).  
CC FT CONFLICT 145 145 R -> A (IN REF. 1).  
CC FT CONFLICT 149 149 D -> G (IN REF. 1).  
CC FT CONFLICT 222 222 M -> V (IN REF. 1).  
CC FT CONFLICT 435 435 A -> H (IN REF. 1).  
CC FT CONFLICT 465 465 V -> A (IN REF. 1).  
CC FT CONFLICT 643 644 MS -> IT (IN REF. 1).  
CC FT CONFLICT 649 649 A -> P (IN REF. 1).  
CC FT CONFLICT 734 734 E -> D (IN REF. 1).  
CC FT CONFLICT 765 765 D -> G (IN REF. 1).  
CC FT CONFLICT 836 836 V -> I (IN REF. 1).  
CC FT CONFLICT 840 840 D -> G (IN REF. 1).  
CC SQ SEQUENCE 856 AA; 96683 MW; B3902C7D98946AD CRC64;  
Query Match 17.2%; Score 89; DB 1; Length 856;  
Best Local Similarity 38.2%; Pred. No. 18;







GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2004, 18:17:42 ; Search time 27.1059 Seconds  
(without alignments)  
1028.178 Million cell updates/sec

Title: US-09-930-169-2  
Perfect score: 516  
Sequence: 1 MANNDAVKRLKLEQKGAADQ.....NSMVSENVIOSTAVTVSSG 108

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues  
Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_23:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_protist:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virins:\*  
16: sp\_bacteriophage:\*  
17: sp\_archaeal:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	392	76.0	319	11 Q8C2U7	Q8C2U7 mus musculus
2	106	20.5	204	5 Q8BYK9	Q8BYK9 drosophila
3	106	20.5	323	5 Q9V504	Q9V504 drosophila
4	101	19.6	270	17 Q58907	Q58907 pyrococcus
5	96	18.6	163	17 Q58287	Q58287 pyrococcus
6	96	18.6	1900	5 Q8IMX3	Q8IMX3 drosophila
7	96	18.6	2028	5 Q9VCD1	Q9VCD1 drosophila
8	95	18.4	570	2 Q9XDC5	Q9XDC5 streptococcus
9	95	18.4	570	16 Q8NZ44	Q8NZ44 streptococcus
10	94.5	18.3	837	11 Q9JH24	Q9JH24 rat mus musculus
11	94	18.2	213	16 Q9K9R4	Q9K9R4 bacillus
12	94	18.2	609	17 Q8TXA4	Q8TXA4 methanopyru
13	93	18.0	534	2 Q38898	Q38898 streptococcus
14	93	18.0	534	2 Q68165	Q68165 streptococcus
15	92.5	17.9	1024	4 Q14527	Q14527 homo sapien
16	92.5	17.9	1046	4 Q9UDT6	Q9UDT6 homo sapien

17	92	17.8	2385	5 Q96216	Q96216 plasmidium
18	91.5	17.7	576	2 Q9KJ33	Q9KJ33 enterococcus
19	91.5	17.7	806	11 Q8VD04	Q8VD04 mus musculus
20	91	17.6	158	17 Q8U3W5	Q8U3W5 pyrococcus
21	91	17.6	2760	5 Q815Y2	Q815Y2 plasmidium
22	90	17.4	987	11 Q8CHF1	Q8CHF1 mus musculus
23	89.5	17.3	189	11 Q9UN83	Q9UN83 rat mus musculus
24	89	17.2	160	17 Q9UYL5	Q9UYL5 pyrococcus
25	89	17.2	816	16 Q8R7S1	Q8R7S1 thermococcus
26	89	17.2	1177	17 Q9VIR8	Q9VIR8 pyrococcus
27	88.5	17.2	673	3 Q74228	Q74228 emericella
28	88.5	17.2	718	13 Q9W6G9	Q9W6G9 xenopus lae
29	88	17.1	148	11 Q8VE59	Q8VE59 mus musculus
30	88	17.1	291	17 Q8TVS0	Q8TVS0 methanopyru
31	88	17.1	319	11 Q9EGX3	Q9EGX3 rat mus musculus
32	87.5	17.0	315	11 Q9DDQ4	Q9DDQ4 mus musculus
33	87.5	17.0	345	4 Q8TAR7	Q8TAR7 homo sapien
34	87.5	17.0	345	4 Q9UH50	Q9UH50 homo sapien
35	87.5	17.0	345	4 Q9GZM8	Q9GZM8 homo sapien
36	87.5	17.0	345	6 Q46480	Q46480 oryctolagus
37	87.5	17.0	345	11 Q9ERR1	Q9ERR1 mus musculus
38	87.5	17.0	345	11 Q9ERT6	Q9ERT6 mus musculus
39	87.5	17.0	1631	11 Q8CHG3	Q8CHG3 drosophila
40	87	16.9	639	5 Q8MRC0	Q8MRC0 drosophila
41	87	16.9	1395	5 Q9WIR4	Q9WIR4 drosophila
42	87	16.9	2910	5 Q8IBY8	Q8IBY8 plasmidium
43	86.5	16.8	192	16 Q8R9Z3	Q8R9Z3 thermococcus
44	86.5	16.8	548	2 Q9RA74	Q9RA74 streptococcus
45	86.5	16.8	574	4 Q9H810	Q9H810 homo sapien

## ALIGNMENTS

### RESULT 1

ID	Q8C2U7	PRELIMINARY;	PRT;	319 AA.
Q8C2U7	Q8C2U7	PRELIMINARY;	PRT;	319 AA.
DT	01-MAR-2003 (TREMURel. 23, Created)			
DT	01-MAR-2003 (TREMURel. 23, Last sequence update)			
DT	01-MAR-2003 (TREMURel. 23, Last annotation update)			
DE	Endothelial monocyte activating polypeptide 2.			
OS	Mus musculus (Mouse).			
CC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
CC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.			
OK	NCBI TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=NOD; TISSUE=Thymus;			
RX	MEDLINE=2234683; PubMed=12466851;			
RA	The FANTOM Consortium,			
RA	The RIKEN Genome Exploration Research Group Phase I & II Team;			
RT	"Analysis of the mouse transcriptome based on functional annotation of			
RT	60,770 full-length cDNAs."			
RL	Nature 420:563-573(2002).			
DR	EMBL: AK087932; BAC40045.1;			
SD	SEQUENCE 319 AA; 35197 MW; B97788E742BEE6 CRC64;			

Qy	Query Match	76.0%; Score 392; DB 11; Length 319;
Qy	Best Local Similarity	76.6%; Pred. No. 7.9e-22;
Db	Matches	82; Conservative 9; Mismatches 16; Indels 0; Gaps 0;
Qy	1 MANNDAVKRLKLEQKGAADQIIIEYLKQVSLKELKALQTLREKKLRVBNATLKEIE 60	
Db	10 MANNDAVKRLKLEQKGAADQIIIEYLKQVSLKELKALQTLREKKLRVBNATLKEIE 69	
Qy	61 ELKQELIQAEIQNGVQKQIAPSGTPLHANGMVSENVIOSTAVTVSS 107	
Db	70 ELKQELIQAEIQNGVQKQIAPSGTPLHANGMVSENVIOSTAVTVSS 116	
RESULT 2		
Q8BYK9		

```

ID 08SYK9 PRELIMINARY; PRT; 204 AA.
AC 08SYK9;
DT 01-JUN-2002 (T-EMBLrel. 21, Created)
DT 01-JUN-2002 (T-EMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE RE55091P.
GN CG8235.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RA Stapleton M., Brokstein P., Hong L., Agbayan A., Carlson J.,
RA Champagne M., Chavez C., Dorsett V., Dreonek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA Patel S., Phoumenavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celniker S.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY071480; AAL49102.1; -.
SQ SEQUENCE 204 AA; 22902 MW; 3840E6034E8535B6 CRC64;

Query Match 20.5%; Score 106; DB 5; Length 204;
Best Local Similarity 31.6%; Pred. No. 1.1; Mismatches 27; Indels 4; Gaps 1;
Matches 24; Conservative 21;

QY 6 AVKRLKQKGAADQIEYKQVSLKKEKALQATLREKKLRVENAKLKEIEELKOE 65
DB 31 ADLQQLASNNRBAALINSIEAISIGIQ-----QLVERKQGEILKEMALAKEVEALAQ 86
QY 66 LIAQETONGVKQIAPF 81
DB 87 LVQLEIRNGKQIAPF 102
DB 87 LVQLEIRNGKQIAPF 102

RESULT 3
QY 09VS04 PRELIMINARY; PRT; 323 AA.
AC 09VS04;
DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE CG8235 protein.
GN CG8235.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RA MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champagne M., Pfeiffer B.L.G.,
RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayan A., An H.-U., Andrews-Pfannkoch C., Baldwin D.,
RA Baillew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotcher P.,
RA Burks K.C., Busam D.A., Butler H., Cadiou E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Fertiera S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,

```

```

RA Glodex A., Gong F., Gozrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Minkulov G., Miletshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacleb J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirek R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN 12
RP SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Barton J., An H., Baldwin D., Barton J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champagne M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dreonek D., Farfan D.,
RA Ferrera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hoiland D., Howland T.J.,
RA Ibegwan C.M., Houck J., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phoumenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Swirek R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RL "Sequencing of Drosophila melanogaster genome."
RT Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN 13
RP SEQUENCE FROM N.A.
RA Mista S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminler J.S., Prochuk S.E., Smith C.D.,
RA Tuhy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RL "Annotation of Drosophila melanogaster genome."
RT Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN 14
RP SEQUENCE FROM N.A.
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN 15
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AEO01835; AAF59019.2; -.
DR FlyBase; FBgn0033351; CG8235.
DR InterPro; IPR002547; tRNA bind.
DR Pfam; PF01588; tRNA_bind.1.
SQ SEQUENCE 323 AA; 34401 MW; 6EBB6E30115FAE84 CRC64;

Query Match 20.5%; Score 106; DB 5; Length 323;
Best Local Similarity 31.6%; Pred. No. 1.8; Mismatches 27; Indels 4; Gaps 1;
Matches 24; Conservative 21;

QY 6 AVKRLKQKGAADQIEYKQVSLKKEKALQATLREKKLRVENAKLKEIEELKOE 65
DB 31 ADLQQLASNNRBAALINSIEAISIGIQ-----QLVERKQGEILKEMALAKEVEALAQ 86
QY 66 LIAQETONGVKQIAPF 81

```

Db 87 LV0LELRNGKQI01PVP 102

RESULT 4

OS8907 058907 PRELIMINARY; PRT; 270 AA.

AC 058907; (TREMBlrel. 07, Created)

DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)

DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)

DE Hypothetical protein PH1189.

GN PH1189.

OS Pyrococcus horikoshii.

OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;

OC Pyrococcus.

NCBI\_TaxID=53953;

NCBI\_TaxID=53953;

RP SEQUENCE FROM N.A.

RC STRAIN=OT3;

RX MEDLINE=98344137; PubMed=9679194;

RA Kawarabayashi Y., Sawada M., Horikawa H., Hainawa Y., Hino Y., Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y., Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kishida N., Oguchi A., Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K., Masuchi Y., Shizuya H., Kikuchi H.;

RA "Complete sequence and gene organization of the genome of a hyper-thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";

RA DNA Res. 5:55-76(1998).

Query Match 19.6%; Score 101; DB 17; Length 270;

Best Local Similarity 30.0%; Pred. No. 3.5;

Matches 30; Conservative 22; Mismatches 32; Indels 16; Gaps 3;

3 NNDVILKRLLEOK-----GAGADIIIVLYKQVSLKELKAILQATLEKKLVENKX 55

Db 86 NNVALISSEKPEVFKSNBETSKETLLEENKLEQ--IQNTLENNQLKNENKQL 143

OY 56 KKEIEELKQELIOLAEICNGVQKQIAFPSSGTPHANSWSEN 95

Db 144 KRRISDLKHLKBAKAQD-----ISRLQVQINNTREN 176

RESULT 5

OS8287 058287 PRELIMINARY; PRT; 163 AA.

AC 058287; (TREMBlrel. 07, Created)

DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)

DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)

DE Hypothetical protein PH0552.

GN PH0552.

OS Pyrococcus horikoshii.

OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;

OC Pyrococcus.

NCBI\_TaxID=53953;

NCBI\_TaxID=53953;

RP SEQUENCE FROM N.A.

RC STRAIN=OT3;

RX MEDLINE=98344137; PubMed=9679194;

RA Kawarabayashi Y., Sawada M., Horikawa H., Hainawa Y., Hino Y., Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y., Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kishida N., Oguchi A., Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K., Masuchi Y., Shizuya H., Kikuchi H.;

RA "Complete sequence and gene organization of the genome of a hyper-thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";

RA DNA Res. 5:55-76(1998).

DR EMBL; AP000002; BAA29641.1; -

DR InterPro; IPR004883; DUF260.

DR Pfam; PF03195; DUF260; 1.

KM Hypothetical protein; Complete proteome.

SO SEQUENCE 163 AA; 19266 MW; 86C178CF9CF2EBSF CRC64;

Query Match 18.6%; Score 96; DB 17; Length 163;

Best Local Similarity 33.3%; Pred. No. 5;

Matches 24; Conservative 19; Mismatches 21; Indels 8; Gaps 2;

OY 8 LKRLBQKGAADQIIEYKQVSLKELKAILQATLEKKLVENKXKKEIEELKQELI 67

Db 83 IIRLDKTKIKIMMLYEVQSQINPFKEQ-----LSQSSLSSEIQELKQIEELKQEL 135

OY 68 QAEIONGVQKQIA 79

Db 136 -AQVNDIKVLA 146

RESULT 6

OSIMX3 08IMX3 PRELIMINARY; PRT; 1900 AA.

AC 08IMX3; (TREMBlrel. 23, Created)

DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)

DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)

DE CG6129-PA.

GN CG6129-PA.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

NCBI\_TaxID=7227;

NCBI\_TaxID=7227;

RP SEQUENCE FROM N.A.

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celiker S.E., Holt R.A., Evans C.A., Gocayne J.D., Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F., George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N., Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X., Brandon R.C., Rogers Y.H., Blazer R.G., Champe M., Pfeiffer B.D., Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L., Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D., Ballew R.M., Basu A., Baxendale J., Bayraktoglu L., Beasley E.M., Beeson K.Y., Berens P.V., Berman B.P., Bhandari D., Bolintsov S., Borokova D., Botchan M.R., Bouck J., Brodeur P., Brotter P., Burtis K.C., Busam D.A., Butler H., Cadien E., Center A., Chandra I., Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P., de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M., Durbin K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P., Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W., Foeller C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K., Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J., Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibeagwa C., Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A., Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z., Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X., Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D., Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A., Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L., Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M., Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G., Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H., Shie B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T., Spier B.E., Spradling A.C., Stapleton M., Strong R., Sun E., Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X., Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissbach J., Williams S.M., Woodgerg, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J., Yen R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L., Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O., Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RA "The genome sequence of Drosophila melanogaster.";

Science 287:2185-2195(2000).

[2]

SEQUENCE FROM N.A.

RA Ceiniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,

RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,

RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busan D.A.,

RA Carlson J.W., Center A., Champagne M., Davenport L.B., Dietz S.M.,

RA Dodson K., Dorsett V., Doup L.B., Doyle C., Dreenek D., Farfan D.,

RA Fiertera S., Frise E., Galle R.F., Garg N.S., George R.A.,

RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,

RA Idegawa C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,

RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,

RA Pacleb U., Paragas V., Park S., Pfeiffer B., Pfeiffer B.,

RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,

RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,

RA Williams S.M., Zaveri J.S., Smith H.O., Venter C.J., Rubin G.M.,

RT "Sequencing of *Drosophila melanogaster* genome."

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN

RP SEQUENCE FROM N.A.

RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,

RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,

RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Ceiniker S.E.,

RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,

RA Krommiller B., Marshall B., Milburn G., Richter J., Russo S.,

RA Searle S.M., Smith E., Shu S., Smutnick F., Whitfield E.,

RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.,

RT "Annotation of *Drosophila melanogaster* genome."

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN

RP SEQUENCE FROM N.A.

RA Adams M.D., Ceiniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.,

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN

RP SEQUENCE FROM N.A.

RA FlyBase;

RA Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.

RL EMBL: AEO03746; AAI13982.1;

SO SEQUENCE 1900 AA; 215425 MW; 3BD3CE23B03002A CAC64;

Query Match 18.6%; Score 96; DB 5; Length 1900;

Best Local Similarity 26.1%; Pred. No. 54;

Matches 36; Conservative 15; Mismatches 41; Indels 46; Gaps 4;

QY 11 LKQKGAADITETLYKQVVS-----LKKKA-----I 37

DB 679 LEOQRIESDNLINLEKQSDLEYDLDKLLKCDLQEKHEKLSNNSCSTDELKSVQNC 738

QY 38 LQATLEBEKRLRYENAKLKEIEELKQEL-----IOAEIONGVQIAFPSCGPLHANS 90

DB 739 LQGAQERKRLKQSDVQNCIEIGELKELALIDKARLELETDN-----LSAEKXKCLQ 792

QY 91 MSENANIQSTAVTVSSG 108

DB 793 LEKEXILQDLACVTRDRG 810

RESULT 7

QVCD1 PRELIMINARY; PRT; 2028 AA.

AC QVCD1; 01-MAY-2000 (T-EMBLrel. 13, Created)

DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)

DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)

DE CG6129 protein.

GN CG6129.

OS *Drosophila melanogaster* (fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephyroidea; *Drosophilidae*; *Drosophila*.

OX NCBI\_Taxid=7227;

RL [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Berkley;

RA MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Mortan J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champagne M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baker E.G., Helt G., Nelson C.R., Miles G.L.G.,

RA Abil J.F., Achayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,

RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhanderi D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Broctler P.,

RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodeon K., Doup L.B., Domes M., Dugan-Rocha S., Dunkel B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Fiertera S., Fleischmann W.,

RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glaser K.,

RA Glodok A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,

RA Jatali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Murzy D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacleb U.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spredling A.C., Stapleton M., Strong R., Sun E.,

RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter C.J.,

RT "The genome sequence of *Drosophila melanogaster*."

RL Science 287:2185-2195(2000).

RN

RP SEQUENCE FROM N.A.

RA Ceiniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,

RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,

RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busan D.A.,

RA Carlson J.W., Center A., Champagne M., Davenport L.B., Dietz S.M.,

RA Dodson K., Dorsett V., Doup L.B., Doyle C., Dreenek D., Farfan D.,

RA Fiertera S., Frise E., Galle R.F., Garg N.S., George R.A.,

RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,

RA Idegawa C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,

RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,

RA Pacleb U., Paragas V., Park S., Pfeiffer B., Pfeiffer B.,

RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,

RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,

RA Williams S.M., Zaveri J.S., Smith H.O., Venter C.J., Rubin G.M.,

RT "Sequencing of *Drosophila melanogaster* genome."

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN

RP SEQUENCE FROM N.A.

RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,

RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,

RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Ceiniker S.E.,

RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,

RA Krommiller B., Marshall B., Milburn G., Richter J., Russo S.,

RA Searle S.M., Smith E., Shu S., Smutnick F., Whitfield E.,

RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.,

RT "Annotation of *Drosophila melanogaster* genome."

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN

RP SEQUENCE FROM N.A.

RA Adams M.D., Ceiniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.,

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN

RP SEQUENCE FROM N.A.



DE GRIP-associated protein 1 long form.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 RN NCBI\_TaxID=10116;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sprague-Dawley;  
 RX MEDLINE=20353053; PubMed=10896157;  
 RA Ye B., Liao D., Zhang X., Zhang P., Dong H., Hugenir R.L.;  
 RT "GRASP-1: A Neuronal RasGEF Associated with the AMPA Receptor/GRIP  
 Complex.";  
 RL Neuron 26:603-617(2000).  
 DR EMBL: AF274057; AAF82298.1;  
 SQ SEQUENCE 837 AA; 96074 MW; A746AEAFD09D3AD2 CRC64;

Query Match 18.3%; Score 94.5; DB 11; Length 837;  
 Best Local Similarity 27.8%; Pred. No. 32;  
 Matches 32; Conservative 20; Mismatches 48; Indels 15; Gaps 3;

QY 1 MANNDAVLKRLBOKGADQIIEYLKQVS-----LLKKAIIQATLR-BEKK 47  
 DB 31 LRKNQVLSLRQVAVLDR--EFSKQKALSKSKAQEVEVLSEKEMIQAKLHSGEED 88  
 QY 48 LRVENAKLKEIEBELKQELIOAEIQNGVKQIAPPSGTPHLANGSVSENVIOSTAV 102  
 DB 89 FRIQNSTLMAFESLCSQLQLELNNQLKEGVPGAGPRVDELLRLQLENTAL 143

RESULT 11  
 ID 09K9R4 PRELIMINARY; PRT; 213 AA.  
 AC 09K9R4;  
 DT 01-OCT-2000 (TRENBLrel. 15, Created)  
 DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)  
 DE Hypothetical protein BH2581.  
 GN BH2581.  
 OS Bacillus halodurans.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=86665;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C-125 / JCM 9153;  
 RX MEDLINE=20512582; PubMed=11058132;  
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Maui N.,  
 RA Fujii F., Hiramata C., Nakamura Y., Ogasawara N., Kuhara S.,  
 RA Horikoshi K.;  
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus  
 RT halodurans and genomic sequence comparison with Bacillus subtilis.";  
 RL Nucleic Acids Res. 28:4317-4331(2000).  
 DR EMBL: AP001516; BAB06300.1;  
 DR InterPro: IPR001005; MyD\_DNA\_binding.  
 DR SMART: SM00717; SANT; 1.  
 DR PROSITE: PS00037; MYB 1; 1.  
 DR PROSITE: PS50090; MYB 3; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 213 AA; 24524 MW; D2C04E7F43673BB6 CRC64;

Query Match 18.2%; Score 94; DB 16; Length 213;  
 Best Local Similarity 36.8%; Pred. No. 9.2;  
 Matches 25; Conservative 16; Mismatches 15; Indels 12; Gaps 3;

QY 7 VLKRLBOKGA--ADQIIEYLKQVS-----KKKAIQATLRBEKKLRVENAKL 57  
 DB 115 VSSGVNDNGSGITWEKVIAPLKNQOTLLGDDRLQKSK---QQLDENETLRKNNAL 171  
 QY 58 EIEELKOE 65  
 DB 172 EIKKROE 179

RESULT 12

Q8TXA4  
 ID 08TXA4 PRELIMINARY; PRT; 609 AA.  
 AC 08TXA4;  
 DT 01-JUN-2002 (TRENBLrel. 21, Created)  
 DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)  
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)  
 DE Uncharacterized protein.  
 GN MK0771.  
 OS Methanopyrus kandleri.  
 OC Archaea; Euryarchaeota; Methanopyrii; Methanopyrales; Methanopyracaceae;  
 OC Methanopyrus.  
 OX NCBI\_TaxID=2320;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AV19 / DSM 6324 / JCM 9639;  
 RX MEDLINE=21927647; PubMed=11930014;  
 RA Slesarev A.I., Mezhevaya K.V., Makarova K.S., Polushin N.N.,  
 RA Shcherbina O.V., Shakhova V.V., Belova G.I., Aravind L.,  
 RA Natale D.A., Rogozin I.B., Tatunov R.L., Wolf Y.I., Stetter K.O.,  
 RA Malykh A.G., Koonin E.V., Kozlovskii S.A.;  
 RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19  
 RT and monophyly of archaeal methanogens.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).  
 DR EMBL: AB010369; AA01985.1;  
 DR HSSP: P04268; IIC2.  
 DR InterPro: IPR002017; Spectrin.  
 KW Complete proteome.  
 SQ SEQUENCE 609 AA; 69552 MW; AB10C9780DC5AD78 CRC64;

Query Match 18.2%; Score 94; DB 17; Length 609;  
 Best Local Similarity 32.6%; Pred. No. 25;  
 Matches 31; Conservative 21; Mismatches 23; Indels 20; Gaps 4;

QY 1 MANNDAVLKRLBOK-----GAEDQIIEYLKQVSLLKKAIIQATLRBEKK----- 47  
 DB 209 LAEN--LKKLKEKNEIKERDRLEETKYEKGLKQDLAKQSKLKEYSRDLANV 265  
 QY 48 --LRVENAKLKEIEBELKQEL--IOAEIQNGVKQI 78  
 DB 266 EALRNENELRKIKDKLSLNSLQKLDREKRL 300

RESULT 13  
 ID 033898 PRELIMINARY; PRT; 534 AA.  
 AC 033898;  
 DT 01-JAN-1998 (TRENBLrel. 05, Created)  
 DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)  
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)  
 DE M-protein.  
 GN SEM.  
 OS Streptococcus equi.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1336;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RA Timoney J.F., Artushin S.C.;  
 RT "The M-protein of Streptococcus equi.";  
 RL Submitted (OCT-1996) to the EMBL/Genbank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE CELL WALL PEPTIDOGLYCAN BY  
 CC AN AMIDE BOND (BY SIMILARITY).  
 DR EMBL: U73162; AAB71984.1;  
 DR InterPro: IPR005877; Gpos\_YsIRK.  
 DR InterPro: IPR001899; Gram\_pos\_anchor.  
 DR InterPro: IPR006192; LPXTG.  
 DR Pfam: PF00746; Gram\_pos\_anchor; 1.  
 DR Pfam: PF04650; YsIRK\_signal; 1.  
 DR PRINTS: PR00015; GPOSANCHOR.  
 DR TIGRPFAM: TIGR01167; LPXTG\_anchor; 1.  
 DR TIGRPFAM: TIGR01168; YsIRK\_signal; 1.  
 DR PROSITE: PS50847; GRAM\_pos\_ANCHORING; 1.  
 KW Cell wall; Peptidoglycan-anchor.

SO SEQUENCE 534 AA; 58213 MW; 654A4AB90E282801 CRC64;

Query Match 18.0%; Score 93; DB 2; Length 534;  
Best Local Similarity 29.2%; Pred. No. 27;  
Matches 40; Conservative 16; Mismatches 51; Indels 30; Gaps 4;

QY 2 ANNDVLRLEKLEKGAADQIIEYLKQV-----SILKEKAILQATLREKKLRVE-- 51  
DB 385 AASDAKVALEKEVEAAKAEVADLKQAQAKKEELEAVKEKEALEAKIEELKKAHAHEL 444

QY 52 -----NAKKKEIEELKQEL---IQAEIQGVQIAPSGTPIHANSM---- 91  
DB 445 SKLKEMLEKKDHANADIQAEINRLKQELADRIKSLISQGRASQTNPGTTAKAQLPSTG 504

QY 92 VSENVIOSTAVTVSSG 108  
DB 505 ESANPFTIALTVIAG 521

RESULT 14  
ID 068165 PRELIMINARY; PRT; 534 AA.  
AC 068165.

DT 01-AUG-1998 (TRENBLrel. 07, Created)  
DT 01-AUG-1998 (TRENBLrel. 07, Last sequence update)  
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)  
DE Fibrinogen-binding protein.  
GN FBP.

OS Streptococcus equi.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.

OX NCBI\_TaxID=1336;  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN=TW;  
RA MEDLJB=96240236; PubMed=9579073;

RT "Affinity purification and characterization of a fibrinogen-binding  
RT protein complex which protects mice against lethal challenge with  
RT Streptococcus equi subsp. equi";  
RL Microbiology 144:993-1003(1998).  
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE CELL WALL PEPTIDOGLYCAN BY  
CC AN AMIDE BOND (BY SIMILARITY).  
CC EMBL; AF012927; AAC38445.1; -

DR InterPro; IPR005877; Gpos\_YsIRK.  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR InterPro; IPR006192; LPXTG.

DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
DR Pfam; PF04650; YsIRK\_signal; 1.  
DR PRINTS; PR00015; GPOSANCHOR.

DR TIGRfams; TIGR01167; LPXTG\_anchor; 1.  
DR TIGRfams; TIGR01168; YsIRK\_signal; 1.  
DR PROSITE; PS50847; GRAM\_POS\_ANCHORING; 1.  
KW Cell wall, peptidoglycan-anchor.

SO SEQUENCE 534 AA; 58344 MW; A8995371273D2B1A CRC64;

Query Match 18.0%; Score 93; DB 2; Length 534;  
Best Local Similarity 29.9%; Pred. No. 27;  
Matches 41; Conservative 15; Mismatches 51; Indels 30; Gaps 5;

QY 2 ANNDVLRLEKLEKGAADQIIEYLKQVSL---LKEKAILQATLREKKLRVE-- 51  
DB 385 AASDAKVALEKEVEAAKAEVADLKQAQAKKEELEAVKEKEALEAKIEELKKAHAHEL 444

QY 52 -----NAKKKEIEELKQEL---IQAEIQGVQIAPSGTPIHANSM---- 91  
DB 445 SKLKEMLEKKDHANADIQAEINRLKQELADRIKSLISQGRASQTNPGTTAKAQLPSTG 504

QY 92 VSENVIOSTAVTVSSG 108  
DB 505 ESANPFTIALTVIAG 521

RESULT 15

ID 014527 PRELIMINARY; PRT; 1024 AA.  
AC 014527;

DT 01-JAN-1998 (TRENBLrel. 05, Created)  
DT 01-OCT-2001 (TRENBLrel. 18, Last sequence update)  
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)  
DE Hypothetical protein KIAA0291 (Fragment).  
GN KIAA0291.

OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI\_TaxID=9606;  
RN [1]

RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RA Ohara O., Nagase T., Ishikawa K., Nakajima D., Ohira M., Seki N.,  
RA Nomura N.;  
RT "Prediction of the coding sequences of unidentified human genes.";  
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL; AB006629; BAA22960.2; -  
DR InterPro; IPR000938; CAP-gly.  
DR Pfam; PF01302; CAP\_GLY; 2.  
DR PROSITE; PS00845; CAP\_GLY\_1; 2.  
DR PROSITE; PS50245; CAP\_GLY\_2; 2.  
KW Hypothetical protein.

SO SEQUENCE 1024 AA; 112981 MW; 12278FC2678B8620 CRC64;

Query Match 17.9%; Score 92.5; DB 4; Length 1024;  
Best Local Similarity 26.9%; Pred. No. 54;  
Matches 25; Conservative 30; Mismatches 31; Indels 7; Gaps 2;

QY 8 LKRLKQGAADQIIEYLKQVSLKKEKAILQATLREKKLRVENATLKEIEELKQELI 67  
DB 713 LKRLDVEYRQQAQAEIPLKEQISLAERKML-----DYRLQRAQKQKQEVESLRKRL 766

QY 68 QAEIQ-NGVQIAPSGTPIHANSVSENVIOS 99  
DB 767 VAENRLQAEALCSQHTMIRSDISEETIRT 799

Search completed: January 2, 2004, 18:22:21  
Job time : 28.1059 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2004, 18:19:07 ; Search time 12.2824 Seconds  
(without alignments)  
372.044 Million cell updates/sec

Title: US-09-930-169-2  
Perfect score: 516  
Sequence: 1 MANDAVLKRLKQKGAADQ.....NSMSENVIOSTAVTVSSG 108

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PTUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	511	99.0	312	US-08-360-821B-36	Sequence 36, Appl
2	500.5	97.0	310	US-08-129-456A-37	Sequence 37, Appl
3	395	76.6	310	US-08-129-456A-36	Sequence 36, Appl
4	395	76.6	310	US-08-705-868-3	Sequence 3, Appl
5	395	76.6	310	US-09-123-615-3	Sequence 3, Appl
6	395	76.6	310	US-08-360-821B-35	Sequence 35, Appl
7	93	18.0	534	US-09-103-664A-2	Sequence 2, Appl
8	90	17.4	956	US-09-914-259-17	Sequence 17, Appl
9	90	17.4	957	US-09-914-259-16	Sequence 16, Appl
10	89.5	17.3	815	US-09-914-259-18	Sequence 18, Appl
11	88	17.1	667	US-09-071-709-9	Sequence 9, Appl
12	87.5	17.0	459	US-09-071-709-1	Sequence 1, Appl
13	84	16.3	741	US-09-854-856-60	Sequence 60, Appl
14	84	16.3	769	US-09-854-856-44	Sequence 44, Appl
15	84	16.3	801	US-09-854-856-28	Sequence 28, Appl
16	84	16.3	829	US-09-854-856-12	Sequence 12, Appl
17	84	16.3	829	US-09-854-856-54	Sequence 54, Appl
18	84	16.3	922	US-09-854-856-38	Sequence 38, Appl
19	84	16.3	954	US-09-854-856-22	Sequence 22, Appl
20	84	16.3	982	US-09-854-856-6	Sequence 6, Appl
21	84	16.3	1911	US-09-854-856-64	Sequence 64, Appl
22	84	16.3	1939	US-09-854-856-48	Sequence 48, Appl
23	84	16.3	1971	US-09-854-856-32	Sequence 32, Appl
24	84	16.3	1999	US-09-854-856-16	Sequence 16, Appl
25	84	16.3	2004	US-09-854-856-58	Sequence 58, Appl
26	84	16.3	2032	US-09-854-856-42	Sequence 42, Appl
27	84	16.3	2048	US-09-854-856-62	Sequence 62, Appl

28	84	16.3	2064	US-09-854-856-26	Sequence 26, Appl
29	84	16.3	2076	US-09-854-856-46	Sequence 46, Appl
30	84	16.3	2092	US-09-854-856-10	Sequence 10, Appl
31	84	16.3	2108	US-09-854-856-30	Sequence 30, Appl
32	84	16.3	2136	US-09-854-856-14	Sequence 14, Appl
33	84	16.3	2141	US-09-854-856-56	Sequence 56, Appl
34	84	16.3	2157	US-09-854-856-42	Sequence 42, Appl
35	84	16.3	2169	US-09-854-856-40	Sequence 40, Appl
36	84	16.3	2185	US-09-854-856-36	Sequence 36, Appl
37	84	16.3	2201	US-09-854-856-24	Sequence 24, Appl
38	84	16.3	2217	US-09-854-856-20	Sequence 20, Appl
39	84	16.3	2229	US-09-854-856-8	Sequence 8, Appl
40	84	16.3	2245	US-09-854-856-4	Sequence 4, Appl
41	84	16.3	2294	US-09-854-856-50	Sequence 50, Appl
42	84	16.3	2322	US-09-854-856-34	Sequence 34, Appl
43	84	16.3	2354	US-09-854-856-18	Sequence 18, Appl
44	84	16.3	2382	US-09-854-856-2	Sequence 2, Appl
45	83	16.1	414	PCT-US93-03077-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1  
US-08-360-821B-36  
Sequence 36, Application US/08360821B  
Patent No. 6228837  
GENERAL INFORMATION:  
APPLICANT: Stern, David M.  
APPLICANT: Clausen, Matthias  
APPLICANT: Kao, Janet  
APPLICANT: Kayton, Mark  
APPLICANT: Libutti, Steven K  
TITLE OF INVENTION: Endothelial Monocyte Activating Polypeptide II: A Mediator Which Activates Host Response  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham, LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.30, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/360,821B  
FILING DATE: 08-OCT-96  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 41735-A-PCT-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-278-0400  
TELEFAX: 212-391-0525  
INFORMATION FOR SEQ ID NO: 36:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 312 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-360-821B-36  
Query Match 99.0%; Score 511; DB 3; Length 312;  
Best Local Similarity 99.1%; Pred. No. 2.4e-45;  
Matches 107; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 MANDAVLKRLKQKGAADQIIIVLKQGVSLKKKALIQATLRREKKLRVENAKLKEIE 60

Db 1 MANDAVLRLKLEKGAADQIIIEYLKQVSLLEKKAIIQATLREKKLRVENAKLKEIE 60  
QY 61 ELKQELIQAIEIONGVQKQIAFPSSGTPHLANSMSSENVIOSTAVTTVSSG 108  
Db 61 ELKQELIQAIEIONGVQKQIIPPSGTPHLANSMSSENVIOSTAVTTVSSG 108

## RESULT 2

US-08-129-456A-37  
; Sequence 37, Application US/08129456A  
; Patent No. 5641867  
; GENERAL INFORMATION:  
; APPLICANT: Stern, David M.  
; APPLICANT: Claus, Matthias  
; APPLICANT: Kao, Janet  
; APPLICANT: Kayton, Mark  
; APPLICANT: Libutti, Steven K.  
; TITLE OF INVENTION: Endothelial-Monocyte Activating  
; TITLE OF INVENTION: Polypeptide II: A Mediator Which  
; TITLE OF INVENTION: Activates Host Response  
; NUMBER OF SEQUENCES: 37  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Cooper & Dunham LLP  
; STREET: 1185 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0 Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/129,456A  
; FILING DATE: 29-SEP-1993  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: White, John P.  
; REGISTRATION NUMBER: 28,678  
; REFERENCE/DOCKET NUMBER: 41735  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212 278 0400  
; TELEFAX: 212 391 0526  
; INFORMATION FOR SEQ ID NO: 37:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 310 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-129-456A-37

Query Match 97.0%; Score 500.5; DB 1; Length 310;  
Best Local Similarity 98.2%; Pred. No. 2.9e-44;  
Matches 107; Conservative 0; Mismatches 1; Indels 1; Gaps 1;  
QY 1 MANDAVLRLKLEKGAADQIIIEYLKQVSLLEKKAIIQATLREKKLRVENAKLKEIE 60  
Db 1 MANDAVLRLKLEKGAADQIIIEYLKQVSLLEKKAIIQATLREKKLRVENAKLKEIE 60  
QY 61 ELKQELIQAIEIONGVQKQIAFPSSGTPHLANSMSSENVIOSTAVTTVSSG 108  
Db 61 ELKQELIQAIEIONGVQKQIIPPSGTPHLANSMSSENVIOSTAVTTVSSG 108

## RESULT 3

US-08-129-456A-36  
; Sequence 36, Application US/08129456A  
; Patent No. 5641867  
; GENERAL INFORMATION:  
; APPLICANT: Stern, David M.  
; APPLICANT: Claus, Matthias

APPLICANT: Kao, Janet  
APPLICANT: Kayton, Mark  
APPLICANT: Libutti, Steven K.  
TITLE OF INVENTION: Endothelial-Monocyte Activating  
TITLE OF INVENTION: Polypeptide II: A Mediator Which  
TITLE OF INVENTION: Activates Host Response  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0 Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/129,456A  
FILING DATE: 29-SEP-1993  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 41735  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212 278 0400  
TELEFAX: 212 391 0526  
INFORMATION FOR SEQ ID NO: 36:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 310 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-129-456A-36

Query Match 76.6%; Score 395; DB 1; Length 310;  
Best Local Similarity 77.6%; Pred. No. 2.6e-33;  
Matches 83; Conservative 8; Mismatches 16; Indels 0; Gaps 0;  
QY 1 MANDAVLRLKLEKGAADQIIIEYLKQVSLLEKKAIIQATLREKKLRVENAKLKEIE 60  
Db 1 MANDAVLRLKLEKGAADQIIIEYLKQVSLLEKKAIIQATLREKKLRVENAKLKEIE 60  
QY 61 ELKQELIQAIEIONGVQKQIAFPSSGTPHLANSMSSENVIOSTAVTTVSS 107  
Db 61 ELKQELIQAIEIONGVQKQIIPPSGTPHLANSMSSENVIOSTAVTTVSS 107

## RESULT 4

US-08-705-868-3  
; Sequence 3, Application US/08705868  
; Patent No. 5885798  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga  
; APPLICANT: Coleman, Roger  
; APPLICANT: Au-Young, Janice  
; APPLICANT: Murty, Lynn E.  
; TITLE OF INVENTION: NOVEL MONOCYTE ACTIVATING CYTOKINE  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: U.S.  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/705,868  
FILING DATE: Filed Herewith  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0117 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 310 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 498910  
US-08-705-868-3

Query Match 76.6%; Score 395; DB 2; Length 310;  
Best Local Similarity 77.6%; Pred. No. 2.6e-33;  
Matches 83; Conservative 8; Mismatches 16; Indels 0; Gaps 0;

Qy 1 MANDAVLKRLKQKGAEDQIIIEYLKQOVSLKKEKALIQATLREKKLRVENAKLKEIE 60  
Db 1 MANDAVLKRLKQKGAEDQIIIEYLKQOVSLKKEKALIQATLREKKLRVENAKLKEIE 60  
Qy 61 ELKQELIAEIQNGVKQIAFPSTGPLHANSWSENVIOSTAVTVSS 107  
Db 61 ELKQELIAEIQNGVKQIAFPSTGPLHANSWSENVIOSTAVTVSS 107

RESULT 5  
US-09-123-615-3  
Sequence 3, Application US/09123615  
Patent No. 6090377  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Coleman, Roger  
APPLICANT: Au-Young, Janice  
APPLICANT: Murry, Lynn E.  
TITLE OF INVENTION: NOVEL MONOCYTE ACTIVATING CYTOKINE  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: U.S.  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/123,615  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/705,868  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0117 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:  
LENGTH: 310 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 498910  
US-09-123-615-3

Query Match 76.6%; Score 395; DB 3; Length 310;  
Best Local Similarity 77.6%; Pred. No. 2.6e-33;  
Matches 83; Conservative 8; Mismatches 16; Indels 0; Gaps 0;

Qy 1 MANDAVLKRLKQKGAEDQIIIEYLKQOVSLKKEKALIQATLREKKLRVENAKLKEIE 60  
Db 1 MANDAVLKRLKQKGAEDQIIIEYLKQOVSLKKEKALIQATLREKKLRVENAKLKEIE 60  
Qy 61 ELKQELIAEIQNGVKQIAFPSTGPLHANSWSENVIOSTAVTVSS 107  
Db 61 ELKQELIAEIQNGVKQIAFPSTGPLHANSWSENVIOSTAVTVSS 107

RESULT 6  
US-08-360-821B-35  
Sequence 35, Application US/08360821B  
Patent No. 6228837

GENERAL INFORMATION:  
APPLICANT: Stern, David M.  
APPLICANT: Klaus, Matthias  
APPLICANT: Kao, Janet  
APPLICANT: Libutti, Steven K  
TITLE OF INVENTION: Endothelial Monocyte Activating  
PEPTIDE II: A Mediator Which Activates Host Response  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham, LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.30, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/360,821B  
FILING DATE: 08-OCT-96  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 41735-A-PCT-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-391-0525  
TELEFAX: 212-278-0400  
INFORMATION FOR SEQ ID NO: 35:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 310 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-360-821B-35

Query Match 76.6%; Score 395; DB 3; Length 310;  
Best Local Similarity 77.6%; Pred. No. 2.6e-33;  
Matches 83; Conservative 8; Mismatches 16; Indels 0; Gaps 0;

QY 1 MANDAVLRLKLEQKGAADQIIIEYLKQVSLKKAIIQATLREKKLRVENAKLKEIE 60  
DB 1 MATNDVAVLRLKLEQKGAADQIIIEYLKQVSLKKAIIQATLREKKLRVENAKLKEIE 60  
QY 61 ELKQELIQAETIONGVKQIAFPSSGTPPLANSVSENVIOSTAVTTVS 107  
DB 61 ELKQELIQAETIONGVKQIAFPSSGTPPLANSVSENVIOSTAVTTVS 107

RESULT 7  
US-09-103-664A-2  
; Sequence 2, Application US/09103664A  
; Patent No. 6458358  
; GENERAL INFORMATION:  
; APPLICANT: University of Kentucky Research Foundation  
; APPLICANT: Timoney, John  
; APPLICANT: Artushin, Sergey  
; TITLE OF INVENTION: Compounds Encoding the Protective M-like Protein of Streptococcus  
; FILE REFERENCE: 50229-212  
; CURRENT APPLICATION NUMBER: US/09/103,664A  
; CURRENT FILING DATE: 1998-06-23  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 2  
; LENGTH: 534  
; TYPE: PRT  
; ORGANISM: Streptococcus equi  
US-09-103-664A-2

Query Match 18.0%; Score 93; DB 4; Length 534;  
Best Local Similarity 29.2%; Pred. No. 0.11;  
Matches 40; Conservative 16; Mismatches 51; Indels 30; Gaps 4;

QY 2 ANDAVLRLKLEQKGAADQIIIEYLKQVSLKKAIIQATLREKKLRVENAKLKEIE 51  
DB 385 AASDAVAVLRLKLEQKGAADQIIIEYLKQVSLKKAIIQATLREKKLRVENAKLKEIE 444

QY 52 NAKLKEIEELKQELIQAETIONGVKQIAFPSSGTPPLANSVSENVIOSTAVTTVS 91  
DB 445 SKLKEMLKQELIQAETIONGVKQIAFPSSGTPPLANSVSENVIOSTAVTTVS 504

QY 92 VSENVIOSTAVTTVS 108  
DB 505 ESANPFTIALTYIAG 521

RESULT 8  
US-09-914-259-17  
; Sequence 17, Application US/09914259  
; Patent No. 6495336  
; GENERAL INFORMATION:  
; APPLICANT: Makowski, Lee  
; APPLICANT: Hyman, Paul  
; APPLICANT: Williams, Mark  
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES  
; FILE REFERENCE: 8471-010-999  
; CURRENT APPLICATION NUMBER: US/09/914,259  
; CURRENT FILING DATE: 2000-11-21  
; NUMBER OF SEQ ID NOS: 180  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 17  
; LENGTH: 956  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-09-914-259-17

Query Match 17.4%; Score 90; DB 4; Length 956;  
Best Local Similarity 26.1%; Pred. No. 0.47;  
Matches 29; Conservative 25; Mismatches 43; Indels 14; Gaps 4;  
QY 6 AVLRLKLEQKGAADQIIIEYLKQVSLKKAIIQATLREKKLRVENAKLKEIE 55  
DB 6 AVLRLKLEQKGAADQIIIEYLKQVSLKKAIIQATLREKKLRVENAKLKEIE 55

DB 422 SLVRQDDKDEINQSLAEKIQQ--MLDDELASTRRDYEKIQEELTRIQIENEA 479  
QY 56 KKEIEELKQELIQAETIONGVKQIAFPSSGTPPLANSVSENVIOSTAVTTVS 106  
DB 480 KQEVKEVLALELAVNVYDQKQEVEDKT--RANBQTLDELAKQTTTLTTT 528

RESULT 9  
US-09-914-259-16  
; Sequence 16, Application US/09914259  
; Patent No. 6495336  
; GENERAL INFORMATION:  
; APPLICANT: Makowski, Lee  
; APPLICANT: Hyman, Paul  
; APPLICANT: Williams, Mark  
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES  
; FILE REFERENCE: 8471-010-999  
; CURRENT APPLICATION NUMBER: US/09/914,259  
; CURRENT FILING DATE: 2000-11-21  
; NUMBER OF SEQ ID NOS: 180  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 16  
; LENGTH: 957  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-914-259-16

Query Match 17.4%; Score 90; DB 4; Length 957;  
Best Local Similarity 26.1%; Pred. No. 0.47;  
Matches 29; Conservative 25; Mismatches 43; Indels 14; Gaps 4;

QY 6 AVLRLKLEQKGAADQIIIEYLKQVSLKKAIIQATLREKKLRVENAKLKEIE 55  
DB 423 SLVRQDDKDEINQSLAEKIQQ--MLDDELASTRRDYEKIQEELTRIQIENEA 480

QY 56 KKEIEELKQELIQAETIONGVKQIAFPSSGTPPLANSVSENVIOSTAVTTVS 106  
DB 481 KQEVKEVLALELAVNVYDQKQEVEDKT--RANBQTLDELAKQTTTLTTT 529

RESULT 10  
US-09-914-259-18  
; Sequence 18, Application US/09914259  
; Patent No. 6495336  
; GENERAL INFORMATION:  
; APPLICANT: Makowski, Lee  
; APPLICANT: Hyman, Paul  
; APPLICANT: Williams, Mark  
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES  
; FILE REFERENCE: 8471-010-999  
; CURRENT APPLICATION NUMBER: US/09/914,259  
; CURRENT FILING DATE: 2000-11-21  
; NUMBER OF SEQ ID NOS: 180  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 18  
; LENGTH: 915  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans  
US-09-914-259-18

Query Match 17.3%; Score 89.5; DB 4; Length 815;  
Best Local Similarity 36.0%; Pred. No. 0.43;  
Matches 27; Conservative 22; Mismatches 17; Indels 9; Gaps 5;

QY 9 KRLKLEQKGAADQIIIEYLKQVSLKKAIIQATLREKKLRVENAKLKEIEELKQ 64  
DB 440 QQLDEKDDIQAETIONGVKQIAFPSSGTPPLANSVSENVIOSTAVTTVS 496  
QY 65 ELIQAETIONGVKQIA 79  
DB 497 E--GKEMVTRALEBIA 509

```
RESULT 11
US-09-071-709-9
; Sequence 9, Application US/09071709
; Patent No. 6171790
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Tang, Y. Tom
; APPLICANT: Lal, Preeti
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl J.
; APPLICANT: Patterson, Chandra
; TITLE OF INVENTION: HUMAN PROTEASE ASSOCIATED PROTEINS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,709
; FILING DATE: Filed Herewith
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: CERRONE, MICHAEL C.
; REGISTRATION NUMBER: 39,132
; REFERENCE/DOCKET NUMBER: PF-0513 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 667 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-071-709-9

Query Match          17.1%; Score 88; DB 3; Length 667;
Best Local Similarity 29.1%; Pred. No. 0.48;
Matches 32; Conservative 18; Mismatches 50; Indels 10; Gaps 3;

QY 2 ANNDVAVLRLEQKGAADQIIEYLKQOVS-LKEKAILQATLRREKKLRVENAKLKKEIEE 61
DB 332 ANDD-----LERAKRATIVSLDTLKNQAIERNAFLESELDKESLIVSVORLKDARD 386

QY 62 LKQELIQAEIQNGVKQIAFPSCGTPHANSMSENVIQ--STAATVSSG 108
DB 387 LRQELAVRERQOEVTNRKSPSSPTLDCERK--DSAVQASLSLPATPVGKG 434

RESULT 12
US-09-071-709-1
; Sequence 1, Application US/09071709
; Patent No. 6171790
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Tang, Y. Tom
; APPLICANT: Lal, Preeti
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl J.
; APPLICANT: Patterson, Chandra
; TITLE OF INVENTION: HUMAN PROTEASE ASSOCIATED PROTEINS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
```

```
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,709
; FILING DATE: Filed Herewith
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: CERRONE, MICHAEL C.
; REGISTRATION NUMBER: 39,132
; REFERENCE/DOCKET NUMBER: PF-0513 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 459 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: THPINOB01
; CLONE: 031381
US-09-071-709-1

Query Match          17.0%; Score 87.5; DB 3; Length 459;
Best Local Similarity 26.8%; Pred. No. 0.33;
Matches 30; Conservative 24; Mismatches 45; Indels 13; Gaps 4;
```

```
QY 8 LKRLQKGAADQI-----IEYLKQOVS-LKEKAILQATLRREKKLRVENAKLKKEI 59
DB 115 VRELEQANDDLERAKRATIVSLDEPQRLNQAIERNAFLESELDKESLIVSVORLDEA 174

QY 60 ELKQELIQAEIQNGVKQIAFPSCGTPHANSMSENVIQ--STAATVSSG 108
DB 175 RDLRQELAVRERQOEVTNRKSPSSPTLDCERK--DSAVQASLSLPATPVGKG 224

RESULT 13
US-09-854-856-60
; Sequence 60, Application US/09854856
; Patent No. 6541252
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Hibun, Erin
; APPLICANT: Donoho, Gregory
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. 6541252el Human Kinases and Polynucleotides
; FILE REFERENCE: LEX-0178-USA
; CURRENT APPLICATION NUMBER: US/09/854,856
; CURRENT FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: US 60/206,015
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 60
; LENGTH: 741
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(741)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-854-856-60

Query Match          16.3%; Score 84; DB 4; Length 741;
```

Best Local Similarity 34.9%; Pred. No. 1.4;  
Matches 30; Conservative 13; Mismatches 19; Indels 24; Gaps 5;

QY 25 LKQVSLKEKAILQATLREKKL-RVENAKLKEIELEKOEILQAEI-ONGVKQIAPFS 82  
Db 498 IKDVSILIRK-----RQROLVREOEKKQESSLSKQVQSSASQTGIKQLPSAS 550

QY 83 -GTPHLANSWSENVIQSTAVTVSS 107  
Db 551 TGIP-----TASTTSAS 562

RESULT 14  
US-09-854-856-44  
; Sequence 44, Application US/09854856  
; Patent No. 6541252

; GENERAL INFORMATION:  
; APPLICANT: Walke, D. Wade  
; APPLICANT: Hilbun, Erin  
; APPLICANT: Donoho, Gregory  
; APPLICANT: Turner, C. Alexander Jr.  
; TITLE OF INVENTION: No. 6541252el Human Kinases and Polynucleotides  
; TITLE OF INVENTION: Encoding the Same  
; FILE REFERENCE: LEX-0178-USA  
; CURRENT APPLICATION NUMBER: US/09/854,856  
; CURRENT FILING DATE: 2001-05-14  
; PRIOR APPLICATION NUMBER: US 60/206,015  
; PRIOR FILING DATE: 2000-05-19  
; NUMBER OF SEQ ID NOS: 64  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 44  
; LENGTH: 769  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (1)...(769)  
; OTHER INFORMATION: Xaa = Any Amino Acid  
US-09-854-856-44

Query Match 16.3%; Score 84; DB 4; Length 769;  
Best Local Similarity 34.9%; Pred. No. 1.5;  
Matches 30; Conservative 13; Mismatches 19; Indels 24; Gaps 5;

QY 25 LKQVSLKEKAILQATLREKKL-RVENAKLKEIELEKOEILQAEI-ONGVKQIAPFS 82  
Db 498 IKDVSILIRK-----RQROLVREOEKKQESSLSKQVQSSASQTGIKQLPSAS 550

QY 83 -GTPHLANSWSENVIQSTAVTVSS 107  
Db 551 TGIP-----TASTTSAS 562

RESULT 15  
US-09-854-856-28  
; Sequence 28, Application US/09854856  
; Patent No. 6541252  
; GENERAL INFORMATION:  
; APPLICANT: Walke, D. Wade  
; APPLICANT: Hilbun, Erin  
; APPLICANT: Donoho, Gregory  
; APPLICANT: Turner, C. Alexander Jr.  
; TITLE OF INVENTION: No. 6541252el Human Kinases and Polynucleotides  
; TITLE OF INVENTION: Encoding the Same  
; FILE REFERENCE: LEX-0178-USA  
; CURRENT APPLICATION NUMBER: US/09/854,856  
; CURRENT FILING DATE: 2001-05-14  
; PRIOR APPLICATION NUMBER: US 60/206,015  
; PRIOR FILING DATE: 2000-05-19  
; NUMBER OF SEQ ID NOS: 64  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 28  
; LENGTH: 801

; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (1)...(801)  
; OTHER INFORMATION: Xaa = Any Amino Acid  
US-09-854-856-28

Query Match 16.3%; Score 84; DB 4; Length 801;  
Best Local Similarity 34.9%; Pred. No. 1.6;  
Matches 30; Conservative 13; Mismatches 19; Indels 24; Gaps 5;

QY 25 LKQVSLKEKAILQATLREKKL-RVENAKLKEIELEKOEILQAEI-ONGVKQIAPFS 82  
Db 558 IKDVSILIRK-----RQROLVREOEKKQESSLSKQVQSSASQTGIKQLPSAS 610

QY 83 -GTPHLANSWSENVIQSTAVTVSS 107  
Db 611 TGIP-----TASTTSAS 622

Search completed: January 2, 2004, 18:23:41  
Job time : 13.2824 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2004, 18:22:28 ; Search time 24.142 Seconds  
(without alignments)  
897.605 Million cell updates/sec

Title: US-09-930-169-2

Perfect score: 516  
Sequence: 1 MANDAVLKRLKQKGAADQ.....NSMSENVIQSTAVTVSSG 108

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 733937 seqs, 200641211 residues

Total number of hits satisfying chosen parameters: 733937

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: Published Applications\_AA:\*

1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/PC1\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/PC1US\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09C\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10C\_NEW\_PUB.pep:\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	516	100.0	108	11	US-09-930-169-2
2	516	100.0	147	11	US-09-930-169-1
3	511	99.0	312	10	US-09-851-026-36
4	395	76.6	310	10	US-09-851-026-35
5	93	18.0	1130	12	US-10-369-493-6751
6	92	17.8	1169	12	US-10-369-493-1095
7	90	17.4	956	12	US-10-080-608A-17
8	90	17.4	956	12	US-10-080-608A-16
9	90	17.4	957	12	US-10-080-608A-15
10	90	17.4	957	12	US-10-080-608A-14
11	89.5	17.3	815	12	US-10-080-608A-18
12	89.5	17.3	815	12	US-10-370-685-107
13	89	17.2	856	9	US-09-815-242-11310
14	89	17.2	1177	12	US-10-369-493-21558
15	87.5	17.0	1079	12	US-10-369-493-2116

16	85	16.5	223	10	US-09-925-300-1616	Sequence 1616, Ap
17	85	16.5	373	12	US-10-104-047-1995	Sequence 1995, Ap
18	84.5	16.4	857	12	US-10-369-493-9374	Sequence 9374, Ap
19	84.5	16.4	861	12	US-10-310-154-485	Sequence 485, App
20	84.5	16.4	861	12	US-10-369-493-17514	Sequence 17514, A
21	84.5	16.4	1526	12	US-10-369-493-22704	Sequence 22704, A
22	84.5	16.4	1965	12	US-10-369-493-3279	Sequence 3279, Ap
23	84.5	16.4	2288	12	US-10-369-493-6774	Sequence 6774, Ap
24	84	16.3	336	12	US-09-976-782-33	Sequence 33, Appl
25	84	16.3	2135	12	US-10-288-798-9	Sequence 9, Appl
26	84	16.3	2382	15	US-10-196-935A-2	Sequence 2, Appl
27	83.5	16.2	1884	10	US-09-785-770A-17	Sequence 17, Appl
28	83.5	16.2	1907	10	US-09-785-770A-16	Sequence 16, Appl
29	83	16.1	676	12	US-10-256-250-16	Sequence 16, Appl
30	83	16.1	724	12	US-10-023-634-80	Sequence 80, Appl
31	83	16.1	725	11	US-09-978-009A-47	Sequence 47, Appl
32	83	16.1	725	12	US-10-256-250-14	Sequence 14, Appl
33	83	16.1	725	12	US-10-023-634-78	Sequence 78, Appl
34	83	16.1	725	12	US-10-023-634-79	Sequence 79, Appl
35	83	16.1	1069	15	US-10-146-473-77	Sequence 77, Appl
36	83	16.1	1827	12	US-10-369-493-5368	Sequence 5368, Ap
37	82.5	16.0	210	15	US-10-154-251-67	Sequence 67, Appl
38	82.5	16.0	975	12	US-10-094-749-2035	Sequence 2035, Ap
39	82.5	16.0	1164	12	US-10-369-493-9770	Sequence 9770, Ap
40	82.5	16.0	1957	12	US-10-369-493-2070	Sequence 2070, Ap
41	82.5	16.0	2099	15	US-10-128-714-3290	Sequence 3290, Ap
42	82.5	16.0	2405	15	US-10-128-714-8290	Sequence 8290, Ap
43	82	15.9	166	11	US-09-930-169-3	Sequence 3, Appl
44	82	15.9	1188	12	US-10-369-493-17323	Sequence 17323, A
45	81.5	15.8	758	9	US-09-925-299-859	Sequence 859, App

## ALIGNMENTS

RESULT 1  
US-09-930-169-2  
; Sequence 2, Application US/09930169  
; Publication No. US20030004309A1  
GENERAL INFORMATION:  
APPLICANT: KIM, YOUNGHOON  
TITLE OF INVENTION: IMMUNOLOGICAL ENHANCEMENT AGENT COMPRISING N-TERMINAL PEPTIDE OF  
FILE REFERENCE: 058333/0106  
CURRENT FILING DATE: 2001-08-16  
PRIOR APPLICATION NUMBER: KR 2001-11310  
NUMBER OF SEQ ID NOS: 11  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 2  
LENGTH: 108  
TYPE: PRT  
ORGANISM: Unknown Organism  
FEATURE:  
OTHER INFORMATION: Description of Unknown Organism: Mammalian  
OTHER INFORMATION: protein sequence  
US-09-930-169-2

Query Match 100.0%; Score 516; DB 11; Length 108;  
Best Local Similarity 100.0%; Pred. No. 2e-43;  
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MANDAVLKRLKQKGAADQIIIEYLKQVSLKKEKAILQATLREKKLRVENAKLKEIE 60  
DB 1 MANDAVLKRLKQKGAADQIIIEYLKQVSLKKEKAILQATLREKKLRVENAKLKEIE 60  
QY 61 ELKELIQAEIQNGVKQIAFPGSGTPLHANSWSENVIQSTAVTVSSG 108  
DB 61 ELKELIQAEIQNGVKQIAFPGSGTPLHANSWSENVIQSTAVTVSSG 108

RESULT 2  
US-09-930-169-1  
Sequence 1, Application US/09930169  
Publication No. US20030004309A1  
GENERAL INFORMATION:  
APPLICANT: KIM, SUNGHON  
APPLICANT: KO, YOUNG-GYU  
TITLE OF INVENTION: IMMUNOLOGICAL ENHANCEMENT AGENT COMPRISING N-TERMINAL PEPTIDE OF  
FILE REFERENCE: 058333/0106  
CURRENT APPLICATION NUMBER: US/09/930,169  
CURRENT FILING DATE: 2001-08-16  
PRIOR APPLICATION NUMBER: KR 2001-31310  
PRIOR FILING DATE: 2001-06-05  
NUMBER OF SEQ ID NOS: 11  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1  
LENGTH: 147  
TYPE: PRT  
ORGANISM: Unknown Organism  
FEATURE:  
OTHER INFORMATION: Description of Unknown Organism: Mammalian  
US-09-930-169-1  
Query Match 100.0%; Score 516; DB 11; Length 147;  
Best Local Similarity 100.0%; Pred. No. 2,9e-43;  
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MANNDAVLRLEOKGAADQIIIEYLKQVSLKKEKAILQATLREKKLRVENAKLKEIE 60  
DB 1 MANNDAVLRLEOKGAADQIIIEYLKQVSLKKEKAILQATLREKKLRVENAKLKEIE 60  
QY 61 ELKQELIQAEIQNGVKQIAPFSGTPLHANSWSENVIOSTAVTVSSG 108  
DB 61 ELKQELIQAEIQNGVKQIAPFSGTPLHANSWSENVIOSTAVTVSSG 108  
RESULT 3  
US-09-851-026-36  
Sequence 36, Application US/09851026  
Patent No. US20020160957A1  
GENERAL INFORMATION:  
APPLICANT: Stern, David M.  
Klaus, Matthias  
Kao, Janet  
Kayton, Mark  
Libutti, Steven K  
TITLE OF INVENTION: Endothelial Monocyte Activating  
Polypeptide II: A Mediator Which Activates Host Response  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSES: Cooper & Dunham, LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.30, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/851,026  
FILING DATE: 07-May-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/360,821  
FILING DATE: 08-OCT-96  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678

REFERENCE/DOCKET NUMBER: 41735-A-PCT-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-278-0400  
TELEFAX: 212-391-0525  
INFORMATION FOR SEQ ID NO: 36:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 312 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 36:  
US-09-851-026-36  
Query Match 99.0%; Score 511; DB 10; Length 312;  
Best Local Similarity 99.1%; Pred. No. 2,4e-42;  
Matches 107; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 MANNDAVLRLEOKGAADQIIIEYLKQVSLKKEKAILQATLREKKLRVENAKLKEIE 60  
DB 1 MANNDAVLRLEOKGAADQIIIEYLKQVSLKKEKAILQATLREKKLRVENAKLKEIE 60  
QY 61 ELKQELIQAEIQNGVKQIAPFSGTPLHANSWSENVIOSTAVTVSSG 108  
DB 61 ELKQELIQAEIQNGVKQIAPFSGTPLHANSWSENVIOSTAVTVSSG 108  
RESULT 4  
US-09-851-026-35  
Sequence 35, Application US/09851026  
Patent No. US20020160957A1  
GENERAL INFORMATION:  
APPLICANT: Stern, David M.  
Klaus, Matthias  
Kao, Janet  
Kayton, Mark  
Libutti, Steven K  
TITLE OF INVENTION: Endothelial Monocyte Activating  
Polypeptide II: A Mediator Which Activates Host Response  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSES: Cooper & Dunham, LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.30, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/851,026  
FILING DATE: 07-May-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/360,821  
FILING DATE: 08-OCT-96  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 41735-A-PCT-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-278-0400  
TELEFAX: 212-391-0525  
INFORMATION FOR SEQ ID NO: 35:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 310 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide



SEQUENCE DESCRIPTION: SEQ ID NO: 35;  
US-09-851-026-35

Query Match 76.6%; Score 395; DB 10; Length 310;  
Best Local Similarity 77.6%; Pred. No. 6.1e-31;  
Matches 83; Conservative 8; Mismatches 16; Indels 0; Gaps 0;

QY 1 MANDAVILKRLQKGAADQIIIEYLKQVSLKKEKALIQATLREKKLAVENAKLKEIE 60  
DB 1 MATNDVAVLKRLQKGAADQIIIEYLKQVALLKKEKALIQATLREKKLAVENAKLKEIE 60  
QY 61 ELKQELIQAEIQNGVQKQIAFPSGTPPLHANSVSENVIOSTAVTVSS 107  
DB 61 ELKQELIQAEIQNGVQKQIAFPSGTPPLHANSVSENVIOSTAVTVSS 107

RESULT 5  
US-10-369-493-6751

Sequence 6751, Application US/10369493  
Publication No. US20030233675A1  
GENERAL INFORMATION:

APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xianfeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 6751  
LENGTH: 1130  
TYPE: PRT  
ORGANISM: Caenorhabditis elegans

Query Match 18.0%; Score 93; DB 12; Length 1130;  
Best Local Similarity 24.6%; Pred. No. 1.7;  
Matches 29; Conservative 30; Mismatches 33; Indels 26; Gaps 4;

QY 2 ANNDVILKRLQKGAADQIIIEYLKQVSLKKEKALIQATLREKK-----KLAVENAKL 55  
DB 691 SDKXNLLEELSK---NKNIEHLKQELIQLNEKISTETKDESELEKTLAQLEIDNSK 746  
QY 56 KKEIEEL-----KQELIQAEIQNGVQKQIAFPSGTPPLHANSVSENVIOS 99  
DB 747 SDQIEKHLRVNMDLQMGITKDELVKNNE--IKTISAKTQALLESSTVESSTKIAS 802

RESULT 6  
US-10-369-493-1095

Sequence 1095, Application US/10369493  
Publication No. US20030233675A1  
GENERAL INFORMATION:

APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xianfeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 1095  
LENGTH: 1169

TYPE: PRT  
ORGANISM: Methanococcus jannaschii  
US-10-369-493-1095

Query Match 17.8%; Score 92; DB 12; Length 1169;  
Best Local Similarity 26.6%; Pred. No. 2.2;  
Matches 25; Conservative 24; Mismatches 25; Indels 20; Gaps 4;

QY 3 NNDVILKRLQKGAADQIIIEYLKQVSLKKEKALIQATLREKKLAVENAKLKEIE 58  
DB 281 NN--IINELNKNNE--EVLRLKSLKELEVEINDKVDLSINELKQVELENNKKE 336  
QY 59 IEELKQELIQ-----AEIQNGVQKQIA 80  
DB 337 IKETQKTIENRDSIIIEKQIKIEEKIKNLNY 370

RESULT 7  
US-10-080-608A-17

Sequence 17, Application US/10080608A  
Publication No. US20030198956A1  
GENERAL INFORMATION:

APPLICANT: Makowski, Lee  
APPLICANT: Hyman, Paul  
APPLICANT: Williams, Mark  
TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES  
FILE REFERENCE: 8471-010-999  
CURRENT APPLICATION NUMBER: US/10/080,608A  
CURRENT FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 180  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 17  
LENGTH: 956  
TYPE: PRT  
ORGANISM: Mus musculus

US-10-080-608A-17  
Query Match 17.4%; Score 90; DB 12; Length 956;  
Best Local Similarity 26.1%; Pred. No. 2.7;  
Matches 29; Conservative 25; Mismatches 43; Indels 14; Gaps 4;

QY 6 AVILKRLQKGAADQIIIEYLKQVSLKKEKALIQATLREKK-----LAVENAKL 55  
DB 422 SLVRQLDKQDEINQSQQLAKLKKQ--MLDQBELASTRDYKIQEBLRLQIENEA 479  
QY 56 KKEIEELKQELIQAEIQNGVQKQIAFPSGTPPLHANSVSENVIOSTAVTVSS 106  
DB 480 KDEYKEVILQALELAVNYDQKQSEVEDKT--RANEQLTDELAQTTTLTTT 528

RESULT 8  
US-10-370-685-106

Sequence 106, Application US/10370685  
Publication No. US20030215903A1  
GENERAL INFORMATION:

APPLICANT: Goldberg, Edward  
FILE REFERENCE: NANF.P-004  
TITLE OF INVENTION: Nanostructures Containing PNA Joining and Functional Elements  
CURRENT APPLICATION NUMBER: US/10/370,685  
CURRENT FILING DATE: 2003-02-21  
PRIOR APPLICATION NUMBER: 10/080,608  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 159  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 106  
LENGTH: 956  
TYPE: PRT  
ORGANISM: mouse  
US-10-370-685-106

Query Match 17.4%; Score 90; DB 12; Length 956;  
Best Local Similarity 26.1%; Pred. No. 2.7;

```
Matches 29; Conservative 25; Mismatches 43; Indels 14; Gaps 4;
QY 6 AVLKRLKQKGAAD---QIIEYLKQGVSLKKEKAILQATLREKK-----LRVENAKL 55
Db 422 SLVRQDDKDEIQKSOQLAEKLRKQ--MDDQDELLASTRDYEKIQEELTRIQIENEAA 479
QY 56 KKEIEELKQELLQAEIQNGVKQIAPPSGTPPLHANSWSENVIOSTAVTVTS 106
Db 480 KQEVKEVLAQLAEILAVNYDQKSOEVEDKT--RANEQLTDELAQKTTTLTTT 528

RESULT 9
US-10-080-608A-16
; Sequence 16, Application US/10080608A
; Publication No. US20030198956A1
; GENERAL INFORMATION:
; APPLICANT: Makowski, Lee
; APPLICANT: Hyman, Paul
; APPLICANT: Williams, Mark
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
; FILE REFERENCE: 8471-010-999
; CURRENT APPLICATION NUMBER: US/10/080,608A
; CURRENT FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 957
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-080-608A-16

Query Match 17.4%; Score 90; DB 12; Length 957;
Best Local Similarity 26.1%; Pred. No. 2.7;
Matches 29; Conservative 25; Mismatches 43; Indels 14; Gaps 4;
QY 6 AVLKRLKQKGAAD---QIIEYLKQGVSLKKEKAILQATLREKK-----LRVENAKL 55
Db 423 SLVRQDDKDEIQKSOQLAEKLRKQ--MDDQDELLASTRDYEKIQEELTRIQIENEAA 480
QY 56 KKEIEELKQELLQAEIQNGVKQIAPPSGTPPLHANSWSENVIOSTAVTVTS 106
Db 481 KQEVKEVLAQLAEILAVNYDQKSOEVEDKT--RANEQLTDELAQKTTTLTTT 529

RESULT 10
US-10-370-685-105
; Sequence 105, Application US/10370685
; Publication No. US20030215903A1
; GENERAL INFORMATION:
; APPLICANT: Hyman, Paul
; APPLICANT: Goldberg, Edward
; TITLE OF INVENTION: Nanostructures Containing PNA Joining and Functional Elements
; FILE REFERENCE: NANF-P-004
; CURRENT APPLICATION NUMBER: US/10/370,685
; CURRENT FILING DATE: 2003-02-21
; PRIOR APPLICATION NUMBER: 10/080,608
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 159
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 105
; LENGTH: 957
; TYPE: PRT
; ORGANISM: human
US-10-370-685-105

Query Match 17.4%; Score 90; DB 12; Length 957;
Best Local Similarity 26.1%; Pred. No. 2.7;
Matches 29; Conservative 25; Mismatches 43; Indels 14; Gaps 4;
QY 6 AVLKRLKQKGAAD---QIIEYLKQGVSLKKEKAILQATLREKK-----LRVENAKL 55
Db 423 SLVRQDDKDEIQKSOQLAEKLRKQ--MDDQDELLASTRDYEKIQEELTRIQIENEAA 480
```

```
QY 56 KKEIEELKQELLQAEIQNGVKQIAPPSGTPPLHANSWSENVIOSTAVTVTS 106
Db 481 KQEVKEVLAQLAEILAVNYDQKSOEVEDKT--RANEQLTDELAQKTTTLTTT 529

RESULT 11
US-10-080-608A-18
; Sequence 18, Application US/10080608A
; Publication No. US20030198956A1
; GENERAL INFORMATION:
; APPLICANT: Makowski, Lee
; APPLICANT: Hyman, Paul
; APPLICANT: Williams, Mark
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
; FILE REFERENCE: 8471-010-999
; CURRENT APPLICATION NUMBER: US/10/080,608A
; CURRENT FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 815
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-080-608A-18

Query Match 17.3%; Score 89.5; DB 12; Length 815;
Best Local Similarity 36.0%; Pred. No. 2.4;
Matches 27; Conservative 22; Mismatches 17; Indels 9; Gaps 5;
QY 9 KRLKQKGAADQI---IEYLKQGVSLKKEKAILQATLRE-EKKLRVENAKLKKKEIEELKQ 64
Db 440 QQLDEKDEIQKSOQLAEKLRKQV--LLQBEAL--GTMRENEELIRENNRFOKEADKQ 496
QY 65 ELIQAEIQNGVKQIA 79
Db 497 E--GKEMMTALEIA 509

RESULT 12
US-10-370-685-107
; Sequence 107, Application US/10370685
; Publication No. US20030215903A1
; GENERAL INFORMATION:
; APPLICANT: Hyman, Paul
; APPLICANT: Goldberg, Edward
; TITLE OF INVENTION: Nanostructures Containing PNA Joining and Functional Elements
; FILE REFERENCE: NANF-P-004
; CURRENT APPLICATION NUMBER: US/10/370,685
; CURRENT FILING DATE: 2003-02-21
; PRIOR APPLICATION NUMBER: 10/080,608
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 159
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 107
; LENGTH: 815
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-370-685-107

Query Match 17.3%; Score 89.5; DB 12; Length 815;
Best Local Similarity 36.0%; Pred. No. 2.4;
Matches 27; Conservative 22; Mismatches 17; Indels 9; Gaps 5;
QY 9 KRLKQKGAADQI---IEYLKQGVSLKKEKAILQATLRE-EKKLRVENAKLKKKEIEELKQ 64
Db 440 QQLDEKDEIQKSOQLAEKLRKQV--LLQBEAL--GTMRENEELIRENNRFOKEADKQ 496
QY 65 ELIQAEIQNGVKQIA 79
Db 497 E--GKEMMTALEIA 509

RESULT 13
```

US-09-815-242-11310  
; Sequence 11310, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zykkind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; PRIORITY FILING DATE: 2001-03-21  
; PRIORITY APPLICATION NUMBER: 60/191,078  
; PRIORITY FILING DATE: 2000-03-21  
; PRIORITY APPLICATION NUMBER: 60/206,848  
; PRIORITY FILING DATE: 2000-05-23  
; PRIORITY APPLICATION NUMBER: 60/207,727  
; PRIORITY FILING DATE: 2000-05-26  
; PRIORITY APPLICATION NUMBER: 60/242,578  
; PRIORITY FILING DATE: 2000-10-23  
; PRIORITY APPLICATION NUMBER: 60/253,625  
; PRIORITY FILING DATE: 2000-11-27  
; PRIORITY APPLICATION NUMBER: 60/257,931  
; PRIORITY FILING DATE: 2000-12-22  
; PRIORITY APPLICATION NUMBER: 60/269,308  
; PRIORITY FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11310  
; LENGTH: 856  
; TYPE: PRT  
; ORGANISM: Helicobacter pylori  
US-09-815-242-11310

Query Match 17.2%; Score 89; DB 9; Length 856;  
Best Local Similarity 38.2%; Pred. No. 2.9;  
Matches 21; Conservative 14; Mismatches 20; Indels 0; Gaps 0;

Qy 11 LEQGAEDQIIEYKQVSLKEKALQATLREKKLRVEMAKLKKIEELKOE 65  
Db 428 MEKESNAKQOEILKEKLSLKEKIQLEAFENEKEVFKISRLKQWESLKEK 482

RESULT 14  
US-10-369-493-21558  
; Sequence 21558, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; PRIORITY FILING DATE: 2003-02-28  
; PRIORITY APPLICATION NUMBER: US 60/360,039  
; PRIORITY FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 21558  
; LENGTH: 1177  
; TYPE: PRT  
; ORGANISM: Pyrococcus abyssi  
US-10-369-493-21558

Query Match 17.2%; Score 89; DB 12; Length 1177;

Best Local Similarity 29.3%; Pred. No. 4.4;  
Matches 29; Conservative 20; Mismatches 24; Indels 26; Gaps 3;

Qy 4 NDAVLKRLQKGAADQIEYK-----QVSLKEK-----ALIQ 40  
Db 408 NEADIKRLKLEKKEKRLSSRTILTKKLPQIREEVKLEKKEKKAELSNVENKISSISQR 467

Qy 41 TLREKKLRVEMAKLKK---EIELKQELIQAEIQNGVK 76  
Db 468 RRVKEELKETSLSLQVSSLSLESLERELIKAEQSEVR 506

RESULT 15  
US-10-369-493-2116  
; Sequence 2116, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; PRIORITY FILING DATE: 2003-02-28  
; PRIORITY APPLICATION NUMBER: US 60/360,039  
; PRIORITY FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 2116  
; LENGTH: 1079  
; TYPE: PRT  
; ORGANISM: Schizosaccharomyces pombe  
US-10-369-493-2116

Query Match 17.0%; Score 87.5; DB 12; Length 1079;  
Best Local Similarity 21.2%; Pred. No. 5.5;  
Matches 29; Conservative 34; Mismatches 41; Indels 33; Gaps 4;

Qy 3 NNDVLRLEQKGAEDQIIEYKQVSLKEKAL-----QATLREKKLRVENA 53  
Db 246 NVTALQWLEBKRRAREEQRIRREARIALBEKRLAEVEARKERARLKKEKERKKE 305

Qy 54 KLR-----KEIELKQELIQAEIQNGVKQIARPPSG---TPLHAN----- 89  
Db 306 EMKQGRYLSKKQKQALQRRLOQWLESQVAGLSNGKKQKPYTNKKKSNRSQTS 365

Qy 90 SMVSENVIOSTAVTVTS 106  
Db 366 SISSGILSSPATIS 382

Search completed: January 2, 2004, 18:31:22  
Job time : 25.1412 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2004, 18:20:38 ; Search time 141.882 Seconds  
(without alignments)  
692.625 Million cell updates/sec

Title: US-09-930-169-2  
Perfect score: 516  
Sequence: 1 MANDAVYKRLKRGKGAEDQ.....NSMVSENVIQSTAVTVSSG 108

Scoring table: BLOSUM62  
Gaped 10.0 , Gapext 0.5

Searched: 5728757 seqs, 909918778 residues  
Total number of hits satisfying chosen parameters: 5728757

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending\_Patents\_AA\_Main:\*

1:	/cgn2_6/ptodata/1/paa/PCTUS_COMB.pep.*
2:	/cgn2_6/ptodata/1/paa/US06_COMB.pep.*
3:	/cgn2_6/ptodata/1/paa/US07_COMB.pep.*
4:	/cgn2_6/ptodata/1/paa/US08_COMB.pep.*
5:	/cgn2_6/ptodata/1/paa/US081_COMB.pep.*
6:	/cgn2_6/ptodata/1/paa/US082_COMB.pep.*
7:	/cgn2_6/ptodata/1/paa/US083_COMB.pep.*
8:	/cgn2_6/ptodata/1/paa/US084_COMB.pep.*
9:	/cgn2_6/ptodata/1/paa/US085_COMB.pep.*
10:	/cgn2_6/ptodata/1/paa/US086_COMB.pep.*
11:	/cgn2_6/ptodata/1/paa/US087_COMB.pep.*
12:	/cgn2_6/ptodata/1/paa/US088_COMB.pep.*
13:	/cgn2_6/ptodata/1/paa/US089_COMB.pep.*
14:	/cgn2_6/ptodata/1/paa/US090_COMB.pep.*
15:	/cgn2_6/ptodata/1/paa/US091_COMB.pep.*
16:	/cgn2_6/ptodata/1/paa/US092_COMB.pep.*
17:	/cgn2_6/ptodata/1/paa/US093_COMB.pep.*
18:	/cgn2_6/ptodata/1/paa/US094_COMB.pep.*
19:	/cgn2_6/ptodata/1/paa/US095_COMB.pep.*
20:	/cgn2_6/ptodata/1/paa/US096_COMB.pep.*
21:	/cgn2_6/ptodata/1/paa/US097_COMB.pep.*
22:	/cgn2_6/ptodata/1/paa/US097A_COMB.pep.*
23:	/cgn2_6/ptodata/1/paa/US098_COMB.pep.*
24:	/cgn2_6/ptodata/1/paa/US099_COMB.pep.*
25:	/cgn2_6/ptodata/1/paa/US099A_COMB.pep.*
26:	/cgn2_6/ptodata/1/paa/US100_COMB.pep.*
27:	/cgn2_6/ptodata/1/paa/US101_COMB.pep.*
28:	/cgn2_6/ptodata/1/paa/US102_COMB.pep.*
29:	/cgn2_6/ptodata/1/paa/US103_COMB.pep.*
30:	/cgn2_6/ptodata/1/paa/US104_COMB.pep.*
31:	/cgn2_6/ptodata/1/paa/US106_COMB.pep.*
32:	/cgn2_6/ptodata/1/paa/US60_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	516	100.0	108	24	US-09-930-169-2 Sequence 2, Appli

2	516	100.0	147	24	US-09-930-169-1	Sequence 1, Appli
3	516	100.0	312	1	PCT-US02-35563-47	Sequence 47, Appli
4	516	100.0	312	26	US-10-049-065-2	Sequence 2, Appli
5	511	99.0	151	23	US-09-834-366-16965	Sequence 16965, A
6	511	99.0	151	32	US-60-197-873-16965	Sequence 16965, A
7	511	99.0	203	32	US-60-453-050-11272	Sequence 11272, A
8	511	99.0	203	32	US-60-453-135-11272	Sequence 11272, A
9	511	99.0	203	32	US-60-453-444-5918	Sequence 5918, A
10	511	99.0	203	32	US-60-465-241-5919	Sequence 5919, Ap
11	511	99.0	203	32	US-60-466-412-11272	Sequence 11272, A
12	511	99.0	310	22	US-09-791-537-80174	Sequence 80174, A
13	511	99.0	312	23	US-09-851-026-36	Sequence 36, Appli
14	511	99.0	312	32	US-60-453-050-11270	Sequence 11270, A
15	511	99.0	312	32	US-60-453-135-11271	Sequence 11270, A
16	511	99.0	312	32	US-60-453-135-11271	Sequence 11271, A
17	511	99.0	312	32	US-60-455-444-5917	Sequence 5917, Ap
18	511	99.0	312	32	US-60-455-444-5918	Sequence 5918, Ap
19	511	99.0	312	32	US-60-465-241-5917	Sequence 5917, Ap
20	511	99.0	312	32	US-60-465-241-5918	Sequence 5918, Ap
21	511	99.0	312	32	US-60-466-412-11270	Sequence 11270, A
22	511	99.0	312	32	US-60-466-412-11271	Sequence 11271, A
23	511	99.0	312	24	US-09-949-016-7144	Sequence 7144, Ap
24	511	99.0	341	22	US-09-760-475-2636	Sequence 2636, Ap
25	511	99.0	341	28	US-10-227-425-2636	Sequence 2636, Ap
26	511	99.0	359	22	US-09-791-537-12674	Sequence 12674, A
27	425	82.4	310	17	US-09-304-448-3	Sequence 3, Appli
28	395	76.6	310	22	US-09-791-537-30491	Sequence 30491, A
29	395	76.6	310	23	US-09-851-026-35	Sequence 35, Appli
30	395	76.6	294	20	US-09-614-150-4422	Sequence 4422, Ap
31	106	20.5	294	22	US-09-791-537-53480	Sequence 53480, A
32	106	20.5	294	32	US-60-167-217-4508	Sequence 4508, Ap
33	106	20.5	294	32	US-60-167-217-4508	Sequence 4508, Ap
34	106	20.5	294	32	US-60-191-637-4437	Sequence 4437, Ap
35	100.5	19.5	672	19	US-09-537-043-22	Sequence 22, Appli
36	100.5	19.5	672	24	US-09-913-655-22	Sequence 22, Appli
37	96	18.6	163	22	US-09-791-537-90144	Sequence 90144, A
38	96	18.6	789	18	US-09-489-039A-9210	Sequence 9210, Ap
39	96	18.6	789	30	US-10-446-203-9210	Sequence 9210, Ap
40	96	18.6	1975	20	US-09-614-150-13074	Sequence 13074, A
41	96	18.6	1975	32	US-60-167-217-13120	Sequence 13120, A
42	96	18.6	1975	32	US-60-173-464-10670	Sequence 10670, A
43	96	18.6	1975	32	US-60-191-637-13115	Sequence 13115, A
44	96	18.6	1975	32	US-60-191-681-10316	Sequence 10316, A
45	96	18.6	2080	32	US-60-161-932-2022	Sequence 2022, Ap

ALIGNMENTS

RESULT 1  
US-09-930-169-2  
; Sequence 2, Application US/09930169  
; GENERAL INFORMATION:  
; APPLICANT: KIM, SUNGHOON  
; APPLICANT: KO, YOUNG-GYU  
; TITLE OF INVENTION: IMMUNOLOGICAL ENHANCEMENT AGENT COMPRISING N-TERMINAL PEPTIDE OF  
; FILE REFERENCE: 058333/0106  
; CURRENT FILING DATE: 2001-08-16  
; PRIOR APPLICATION NUMBER: KR 2001-31310  
; PRIOR FILING DATE: 2001-06-05  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 108  
; TYPE: PRT  
; ORGANISM: Unknown Organism  
; FEATURE: Description of Unknown Organism: Mammalian  
; OTHER INFORMATION: Protein sequence  
US-09-930-169-2

Query Match 100.0%; Score 516; DB 24; Length 108;  
Best Local Similarity 100.0%; Pred. No. 2.2e-39;  
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MANDAVLRLKLEQKGAADQIIIEYLKQVSLKEKAILQATLREKKLRVENAKLKEIE 60  
DB 1 MANDAVLRLKLEQKGAADQIIIEYLKQVSLKEKAILQATLREKKLRVENAKLKEIE 60  
QY 61 ELKQELIOAEIONGVQKQIAFPSTGPLHANSMSENVIOSTAVTTVSSG 108  
DB 61 ELKQELIOAEIONGVQKQIAFPSTGPLHANSMSENVIOSTAVTTVSSG 108

## RESULT 2

US-09-930-169-1  
Sequence 1, Application US/09930169

GENERAL INFORMATION:  
APPLICANT: KIM, SUNGHOON  
TITLE OF INVENTION: IMMUNOLOGICAL ENHANCEMENT AGENT COMPRISING N-TERMINAL PEPTIDE OF  
FILE REFERENCE: 058333/0106  
CURRENT APPLICATION NUMBER: US/09/930,169  
PRIOR FILING DATE: 2001-08-16  
PRIOR APPLICATION NUMBER: KR 2001-31310  
NUMBER OF SEQ ID NOS: 11  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1  
LENGTH: 147  
TYPE: PRT  
ORGANISM: Unknown Organism  
FEATURE:  
OTHER INFORMATION: Description of Unknown Organism: Mammalian  
OTHER INFORMATION: protein sequence  
US-09-930-169-1

Query Match 100.0%; Score 516; DB 24; Length 147;  
Best Local Similarity 100.0%; Pred. No. 3.4e-39;  
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MANDAVLRLKLEQKGAADQIIIEYLKQVSLKEKAILQATLREKKLRVENAKLKEIE 60  
DB 1 MANDAVLRLKLEQKGAADQIIIEYLKQVSLKEKAILQATLREKKLRVENAKLKEIE 60  
QY 61 ELKQELIOAEIONGVQKQIAFPSTGPLHANSMSENVIOSTAVTTVSSG 108  
DB 61 ELKQELIOAEIONGVQKQIAFPSTGPLHANSMSENVIOSTAVTTVSSG 108

## RESULT 3

PCT-US02-35563-47  
Sequence 47, Application PC/TUS0235563

GENERAL INFORMATION:  
APPLICANT: Origene Technologies, Inc  
TITLE OF INVENTION: REGULATED PROSTATE CANCER GENES  
FILE REFERENCE: OGT 9U 803 PCT  
CURRENT APPLICATION NUMBER: PCT/US02/35563  
PRIOR FILING DATE: 2002-11-07  
PRIOR APPLICATION NUMBER: 60/331,042  
PRIOR FILING DATE: 2001-11-07  
PRIOR APPLICATION NUMBER: 60/331,041  
PRIOR FILING DATE: 2001-11-07  
PRIOR APPLICATION NUMBER: 60/340,251  
PRIOR FILING DATE: 2001-12-18  
PRIOR APPLICATION NUMBER: 60/344,791  
PRIOR FILING DATE: 2002-01-07  
NUMBER OF SEQ ID NOS: 52  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 47  
LENGTH: 312  
TYPE: PRT  
ORGANISM: Homo sapiens

PCT-US02-35563-47

Query Match 100.0%; Score 516; DB 1; Length 312;  
Best Local Similarity 100.0%; Pred. No. 9.7e-39;  
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MANDAVLRLKLEQKGAADQIIIEYLKQVSLKEKAILQATLREKKLRVENAKLKEIE 60  
DB 1 MANDAVLRLKLEQKGAADQIIIEYLKQVSLKEKAILQATLREKKLRVENAKLKEIE 60  
QY 61 ELKQELIOAEIONGVQKQIAFPSTGPLHANSMSENVIOSTAVTTVSSG 108  
DB 61 ELKQELIOAEIONGVQKQIAFPSTGPLHANSMSENVIOSTAVTTVSSG 108

## RESULT 4

US-10-049-065-2  
Sequence 2, Application US/10049065

GENERAL INFORMATION:  
APPLICANT: KIM, SUNGHOON  
APPLICANT: KO, YOUNG-GYU  
APPLICANT: KIM, YOUNG SOO  
TITLE OF INVENTION: 3 ANTITUMOR THERAPEUTIC AGENT AND THREE DIMENSIONAL  
FILE REFERENCE: 058333/0113  
CURRENT APPLICATION NUMBER: US/10/049,065  
CURRENT FILING DATE: 2002-07-08  
PRIOR APPLICATION NUMBER: PCT/KR00/00630  
PRIOR FILING DATE: 2000-06-14  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 2  
LENGTH: 312  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-049-065-2

Query Match 100.0%; Score 516; DB 26; Length 312;  
Best Local Similarity 100.0%; Pred. No. 9.7e-39;  
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MANDAVLRLKLEQKGAADQIIIEYLKQVSLKEKAILQATLREKKLRVENAKLKEIE 60  
DB 1 MANDAVLRLKLEQKGAADQIIIEYLKQVSLKEKAILQATLREKKLRVENAKLKEIE 60  
QY 61 ELKQELIOAEIONGVQKQIAFPSTGPLHANSMSENVIOSTAVTTVSSG 108  
DB 61 ELKQELIOAEIONGVQKQIAFPSTGPLHANSMSENVIOSTAVTTVSSG 108

## RESULT 5

US-09-834-366-16965  
Sequence 16965, Application US/09834366

GENERAL INFORMATION:  
APPLICANT: Benjamin, Stephanie  
APPLICANT: Tanaka, Hiroaki  
APPLICANT: Dumas Malne Edwards, Jean Baptiste  
APPLICANT: Jobert, Severin  
TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
FILE REFERENCE: 81 US2, REG  
CURRENT APPLICATION NUMBER: US/09/834,366  
CURRENT FILING DATE: 2001-04-13  
PRIOR APPLICATION NUMBER: US 60/197,873  
PRIOR FILING DATE: 2000-04-18  
NUMBER OF SEQ ID NOS: 52153  
SOFTWARE: Patent .pm  
SEQ ID NO 16965  
LENGTH: 151  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-834-366-16965

Query Match 99.0%; Score 511; DB 23; Length 151;  
Best Local Similarity 99.1%; Pred. No. 1e-38;  
Matches 107; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MANDDAVLRLEQKGAEDQIIEYLKQOVSLKKEKALIQATLRREKKLRVENAKLKEIE 60  
Db 25 MANDDAVLRLEQKGAEDQIIEYLKQOVSLKKEKALIQATLRREKKLRVENAKLKEIE 84

Qy 61 ELKQELIOAEIQNGVKQIAPPSGTPPLHANSVSENVIOSTAVTTVSSG 108  
Db 85 ELKQELIOAEIQNGVKQIAPPSGTPPLHANSVSENVIOSTAVTTVSSG 132

RESULT 6  
US-60-197-873-16965  
; Sequence 16965, Application US/60197873  
; GENERAL INFORMATION:  
; APPLICANT: Bejani, Stephane  
; APPLICANT: Tanaka, Hiroaki  
; APPLICANT: Dumas Milne Edwards, Jean Baptiste  
; APPLICANT: Jobert, Severin  
; APPLICANT: Giordano, Jean-Yves  
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
; FILE REFERENCE: 81.US1.PRO  
; CURRENT APPLICATION NUMBER: US/60/197,873  
; CURRENT FILING DATE: 2000-04-18  
; NUMBER OF SEQ ID NOS: 52153  
; SOFTWARE: Patent.pm  
; SEQ ID NO 16965  
; LENGTH: 151  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-60-197-873-16965

Query Match 99.0%; Score 511; DB 32; Length 151;  
Best Local Similarity 99.1%; Pred. No. 1e-38;  
Matches 107; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MANDDAVLRLEQKGAEDQIIEYLKQOVSLKKEKALIQATLRREKKLRVENAKLKEIE 60  
Db 25 MANDDAVLRLEQKGAEDQIIEYLKQOVSLKKEKALIQATLRREKKLRVENAKLKEIE 84

Qy 61 ELKQELIOAEIQNGVKQIAPPSGTPPLHANSVSENVIOSTAVTTVSSG 108  
Db 85 ELKQELIOAEIQNGVKQIAPPSGTPPLHANSVSENVIOSTAVTTVSSG 132

RESULT 7  
US-60-453-050-11272  
; Sequence 11272, Application US/60453050  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele  
; APPLICANT: LUKE, May  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
; TITLE OF INVENTION: STENOSIS, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CLO01457  
; CURRENT APPLICATION NUMBER: US/60/453,050  
; CURRENT FILING DATE: 2003-03-10  
; NUMBER OF SEQ ID NOS: 82762  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11272  
; LENGTH: 203  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-60-453-050-11272

Query Match 99.0%; Score 511; DB 32; Length 203;  
Best Local Similarity 99.1%; Pred. No. 1.5e-38;  
Matches 107; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MANDDAVLRLEQKGAEDQIIEYLKQOVSLKKEKALIQATLRREKKLRVENAKLKEIE 60  
Db 61 ELKQELIOAEIQNGVKQIAPPSGTPPLHANSVSENVIOSTAVTTVSSG 108

Db 1 MANDDAVLRLEQKGAEDQIIEYLKQOVSLKKEKALIQATLRREKKLRVENAKLKEIE 60  
Qy 61 ELKQELIOAEIQNGVKQIAPPSGTPPLHANSVSENVIOSTAVTTVSSG 108  
Db 61 ELKQELIOAEIQNGVKQIAPPSGTPPLHANSVSENVIOSTAVTTVSSG 108

RESULT 8  
US-60-453-135-11272  
; Sequence 11272, Application US/60453135  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele  
; APPLICANT: IAKUBOVA, Olga  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
; TITLE OF INVENTION: MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CLO01456  
; CURRENT APPLICATION NUMBER: US/60/453,135  
; CURRENT FILING DATE: 2003-03-10  
; NUMBER OF SEQ ID NOS: 82762  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11272  
; LENGTH: 203  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-60-453-135-11272

Query Match 99.0%; Score 511; DB 32; Length 203;  
Best Local Similarity 99.1%; Pred. No. 1.5e-38;  
Matches 107; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MANDDAVLRLEQKGAEDQIIEYLKQOVSLKKEKALIQATLRREKKLRVENAKLKEIE 60  
Db 1 MANDDAVLRLEQKGAEDQIIEYLKQOVSLKKEKALIQATLRREKKLRVENAKLKEIE 60

Qy 61 ELKQELIOAEIQNGVKQIAPPSGTPPLHANSVSENVIOSTAVTTVSSG 108  
Db 61 ELKQELIOAEIQNGVKQIAPPSGTPPLHANSVSENVIOSTAVTTVSSG 108

RESULT 9  
US-60-455-444-5919  
; Sequence 5919, Application US/60455444  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele  
; APPLICANT: BEGOVICH, Ann  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
; TITLE OF INVENTION: RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CLO01455  
; CURRENT APPLICATION NUMBER: US/60/455,444  
; CURRENT FILING DATE: 2003-03-18  
; NUMBER OF SEQ ID NOS: 50986  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5919  
; LENGTH: 203  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-60-455-444-5919

Query Match 99.0%; Score 511; DB 32; Length 203;  
Best Local Similarity 99.1%; Pred. No. 1.5e-38;  
Matches 107; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MANDDAVLRLEQKGAEDQIIEYLKQOVSLKKEKALIQATLRREKKLRVENAKLKEIE 60  
Db 1 MANDDAVLRLEQKGAEDQIIEYLKQOVSLKKEKALIQATLRREKKLRVENAKLKEIE 60

Qy 61 ELKQELIOAEIQNGVKQIAPPSGTPPLHANSVSENVIOSTAVTTVSSG 108  
Db 61 ELKQELIOAEIQNGVKQIAPPSGTPPLHANSVSENVIOSTAVTTVSSG 108

RESULT 10  
US-60-465-241-5919

; Sequence 5919, Application US/60465241  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele  
; APPLICANT: BEGOVICH, Ann  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
; FILE REFERENCE: CL001468  
; CURRENT APPLICATION NUMBER: US/60/465,241  
; CURRENT FILING DATE: 2003-04-23  
; NUMBER OF SEQ ID NOS: 258418  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5919  
; LENGTH: 203  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-60-465-241-5919

Query Match 99.0%; Score 511; DB 32; Length 203;  
Best Local Similarity 99.1%; Pred. No. 1.5e-38;  
Matches 107; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MANNDAVLKRLKQKGAADQIIIEYLKQVSLKREKALIQATLREKKLRVENAKLKEIE 60  
DB 1 MANNDAVLKRLKQKGAADQIIIEYLKQVSLKREKALIQATLREKKLRVENAKLKEIE 60  
QY 61 ELKQELIQAEIQNGVKQIAFPSTGPLHANSWSENVIOSTAVTTVSSG 108  
DB 61 ELKQELIQAEIQNGVKQIAFPSTGPLHANSWSENVIOSTAVTTVSSG 108

RESULT 11  
US-60-466-412-11272  
; Sequence 11272, Application US/60466412  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele  
; APPLICANT: IAKOUBOVA, Olga  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
; FILE REFERENCE: CL001466  
; CURRENT APPLICATION NUMBER: US/60/466,412  
; CURRENT FILING DATE: 2003-04-30  
; NUMBER OF SEQ ID NOS: 429241  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11272  
; LENGTH: 203  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-60-466-412-11272

Query Match 99.0%; Score 511; DB 32; Length 203;  
Best Local Similarity 99.1%; Pred. No. 1.5e-38;  
Matches 107; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MANNDAVLKRLKQKGAADQIIIEYLKQVSLKREKALIQATLREKKLRVENAKLKEIE 60  
DB 1 MANNDAVLKRLKQKGAADQIIIEYLKQVSLKREKALIQATLREKKLRVENAKLKEIE 60  
QY 61 ELKQELIQAEIQNGVKQIAFPSTGPLHANSWSENVIOSTAVTTVSSG 108  
DB 61 ELKQELIQAEIQNGVKQIAFPSTGPLHANSWSENVIOSTAVTTVSSG 108

RESULT 12  
US-09-791-537-80174  
; Sequence 80174, Application US/09791537  
; GENERAL INFORMATION:  
; APPLICANT: Biomomix, Inc.  
; APPLICANT: Debe, Derek  
; APPLICANT: Danzer, Joseph  
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMB  
; FILE REFERENCE: 261/210  
; CURRENT APPLICATION NUMBER: US/09/791,537

; CURRENT FILING DATE: 2001-02-22  
; NUMBER OF SEQ ID NOS: 153055  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 80174  
; LENGTH: 310  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-791-537-80174

Query Match 99.0%; Score 511; DB 22; Length 310;  
Best Local Similarity 99.1%; Pred. No. 2.8e-38;  
Matches 107; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MANNDAVLKRLKQKGAADQIIIEYLKQVSLKREKALIQATLREKKLRVENAKLKEIE 60  
DB 1 MANNDAVLKRLKQKGAADQIIIEYLKQVSLKREKALIQATLREKKLRVENAKLKEIE 60  
QY 61 ELKQELIQAEIQNGVKQIAFPSTGPLHANSWSENVIOSTAVTTVSSG 108  
DB 61 ELKQELIQAEIQNGVKQIAFPSTGPLHANSWSENVIOSTAVTTVSSG 108

RESULT 13  
US-09-851-026-36  
; Sequence 36, Application US/09851026  
; GENERAL INFORMATION:  
; APPLICANT: Stern, David M.  
; APPLICANT: Klaus, Matthias  
; Kao, Janet  
; Kayton, Mark  
; Libutti, Steven K  
; TITLE OF INVENTION: Endothelial Monocyte Activating  
; Polypeptide II: A Mediator Which Activates Host Response

NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham, LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA

ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.30, Version #1.30

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/851,026  
FILING DATE: 07-May-2001  
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/360,821  
FILING DATE: 08-OCT-96

ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.

REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 41735-A-PCT-US

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-278-0400

TELEFAX: 212-391-0525

INFORMATION FOR SEQ ID NO: 36:

SEQUENCE CHARACTERISTICS:

LENGTH: 312 amino acids  
TYPE: amino acid  
STRANDEDNESS: single

MOLECULE TYPE: peptide  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 36:

US-09-851-026-36  
Query Match 99.0%; Score 511; DB 23; Length 312;  
Best Local Similarity 99.1%; Pred. No. 2.8e-38;  
Matches 107; Conservative 0; Mismatches 1; Indels 0; Gaps 0;



Search completed: January 2, 2004, 18:29:30  
Job time : 142.882 secs

Qy	1	MANNDAVLKRLBQGAADQIIEYLKQVSLLEKALQATTEEEKKLRVNAKKKEIE	60
Db	1	MANNDAVLKRLBQGAADQIIEYLKQVSLLEKALQATTEEEKKLRVNAKKKEIE	60
Qy			
Db			
Qy	61	ELKEELIQAEQNGVKRIAPPSGPHLANSMVENNIQSTAVTTVSSG	108
Db	61	ELKEELIQAEQNGVKRIAPPSGPHLANSMVENNIQSTAVTTVSSG	108

```

RESULT 14
US-60-453-050-11270
: Sequence 11270, Application US/6045050
: GENERAL INFORMATION:
: APPLICANT: CARGILL, Michele
: APPLICANT: LUKE, May
: TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
: TITLE OF INVENTION: STENOSIS, METHODS OF DETECTION AND USES THEREOF
: FILE REFERENCE: C1001457
: CURRENT APPLICATION NUMBER: US/60/453,050
: CURRENT FILING DATE: 2003-03-10
: NUMBER OF SEQ ID NOS: 82762
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 11270
: LENGTH: 312
: TYPE: FRT
: ORGANISM: Homo sapiens
US-60-453-050-11270

```

Query Match	99.0%	Score 511,	DB 32,	Length 312,
Best Local Similarity	99.1%	Pred. No. 2.8e-18,		
Matches 107;	Conservative	0;	Mismatches 1,	Indels 0;
				Gaps 0;

Qy	Dy	Qy	Dy
1	1	61	60
MANNDAVILKRELOGGAADQII	MANNDAVILKRELOGGAADQII	ELKQELLQAEIÖNGVCKOIA	ELKQELLQAEIÖNGVCKOIA
EYLKQOVSLLEKXKILLOAT	EYLKQOVSLLEKXKILLOAT	PPSGCPPLHANSMBENNY	PPSGCPPLHANSMBENNY
EEKKLEVEYNAKKKEIE	EEKKLEVEYNAKKKEIE	IOSTAVTVTVSSG	IOSTAVTVTVSSG
60	60	108	108
1	1	61	60
MANNDAVILKRELOGGAADQII	MANNDAVILKRELOGGAADQII	ELKQELLQAEIÖNGVCKOIA	ELKQELLQAEIÖNGVCKOIA
EYLKQOVSLLEKXKILLOAT	EYLKQOVSLLEKXKILLOAT	PPSGCPPLHANSMBENNY	PPSGCPPLHANSMBENNY
EEKKLEVEYNAKKKEIE	EEKKLEVEYNAKKKEIE	IOSTAVTVTVSSG	IOSTAVTVTVSSG
60	60	108	108

```

RESULT 15
US-60-453-050-11271
: Sequence 11271, Application US/60453050
: GENERAL INFORMATION:
: APPLICANT: CARCELL, Michèle
: APPLICANT: LUKE, May
: TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
: FILE OF INVENTION: STENOSIS, METHODS OF DETECTION AND USES THEREOF
: FILE REFERENCE: CU001457
: CURRENT APPLICATION NUMBER: US/60/453,050
: CURRENT FILING DATE: 2003-03-10
: NUMBER OF SEQ ID NOS: 82762
: SOFTWARE: FastSeq, for Windows Version 4.0
: SEQ ID NO 11271
: LENGTH: 312
: TYPE: PRT
: ORGANISM: Homo sapiens
US-60-453-050-11271

```

Query Match	99.0%	Score 511;	DB 32;	length 312;
Best Local Similarity	99.1%;	Pred. No. 2.8e-38;		
Matches 107;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;

Qy	1	MANNDAVLKRLKLEGGAGADQIIEYLKQOVSLLAEKKIILQATLREKKKLVENNAKKKEIE	60
Db	1	MANNDAVLKRLKLEGGAGADQIIEYLKQVSLLAEKKIILQATLREKKKLVENNAKKKEIE	60
Qy			
Db	61	ELKKEILLQAEIQNGVCKIAPFSGPLHANSMSVENNIQSTAVTVTSSG	108
Qy			
Db	61	ELKKEILLQAEIQNGVCKIAPFSGPLHANSMSVENNIQSTAVTVTSSG	108

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2004, 18:21:08 ; Search time 12.7059 Seconds  
(without alignments)  
548.910 Million cell updates/sec

Title: US-09-930-169-2

Perfect score: 516  
Sequence: 1 MANDAVLKRLKQKGAADQ.....NSMSENVIOSTAVTVSSG 108

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 320202 seqs, 64577629 residues

Total number of hits satisfying chosen parameters: 320202

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Pending Patents: AA New: \*  
1: /cgn2\_6/ptodata/1/paa/PCT\_NEW\_COMB.pep: \*  
2: /cgn2\_6/ptodata/1/paa/US06\_NEW\_COMB.pep: \*  
3: /cgn2\_6/ptodata/1/paa/US07\_NEW\_COMB.pep: \*  
4: /cgn2\_6/ptodata/1/paa/US08\_NEW\_COMB.pep: \*  
5: /cgn2\_6/ptodata/1/paa/US09\_NEW\_COMB.pep: \*  
6: /cgn2\_6/ptodata/1/paa/US10\_NEW\_COMB.pep: \*  
7: /cgn2\_6/ptodata/1/paa/US60\_NEW\_COMB.pep: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	516	100.0	328	5	US-09-979-932A-606 Sequence 606, App
2	289.5	56.1	123	1	PCT-US03-33947-1087 Sequence 1087, App
3	106	20.5	294	5	US-09-614-150A-4422 Sequence 4422, App
4	96	18.6	1975	5	US-09-614-150A-13074 Sequence 13074, App
5	92.5	17.9	1046	1	PCT-US03-38193-1546 Sequence 1546, App
6	92.5	17.9	1046	6	US-10-723-860-1546 Sequence 1546, App
7	92.5	17.9	1047	6	US-10-258-899A-1169 Sequence 1169, App
8	92.5	17.9	1047	6	US-10-258-899A-1137 Sequence 1137, App
9	90	17.4	911	6	US-10-330-773-712 Sequence 712, App
10	90	17.4	952	6	US-10-330-773-709 Sequence 709, App
11	89.5	17.3	300	6	US-10-679-063-2630 Sequence 2630, App
12	87	16.9	1305	5	US-09-614-150A-31653 Sequence 31653, App
13	86.5	16.8	931	6	US-10-258-899A-3150 Sequence 3150, App
14	86.5	16.8	990	6	US-10-258-899A-1182 Sequence 1182, App
15	86	16.7	680	6	US-10-425-114A-68086 Sequence 68086, App
16	86	16.7	1755	6	US-10-677-466-49 Sequence 469, App
17	85.5	16.6	277	6	US-10-425-114A-51330 Sequence 51330, App
18	85.5	16.6	484	6	US-10-258-899A-1647 Sequence 1647, App
19	85.5	16.6	533	6	US-10-258-899A-3615 Sequence 3615, App
20	85.5	16.6	676	6	US-10-425-114A-71797 Sequence 71797, App
21	85.5	16.6	948	7	US-60-490-890-1613 Sequence 1613, App
22	85	16.5	1066	5	US-09-614-150A-19545 Sequence 19545, App
23	84.5	16.4	217	6	US-10-679-063-3049 Sequence 3049, App
24	84.5	16.4	288	6	US-10-679-063-2467 Sequence 2467, App
25	84.5	16.4	295	6	US-10-425-114A-58037 Sequence 58037, App
26	84	16.3	271	6	US-10-425-114A-66836 Sequence 66836, App

27	84	16.3	317	6	US-10-425-114A-66393 Sequence 66393, App
28	84	16.3	1937	1	PCT-US03-38193-58 Sequence 58, App
29	84	16.3	1937	1	PCT-US03-38193-106 Sequence 106, App
30	84	16.3	1937	6	US-10-723-860-58 Sequence 58, App
31	84	16.3	1937	6	US-10-723-860-106 Sequence 106, App
32	84	16.3	1939	7	US-60-485-101-650 Sequence 650, App
33	84	16.3	1981	1	PCT-US03-30720-1357 Sequence 1357, App
34	84	16.3	2070	1	PCT-US03-19660-46 Sequence 46, App
35	84	16.3	2382	7	US-60-490-890-2015 Sequence 2015, App
36	84	16.3	2388	1	PCT-US03-30720-1356 Sequence 1356, App
37	83.5	16.2	1193	5	US-09-976-858-50 Sequence 50, App
38	83.5	16.2	1193	6	US-10-296-115-1117 Sequence 1117, App
39	83.5	16.2	1591	1	PCT-US02-24483-66 Sequence 66, App
40	83.5	16.2	1767	1	PCT-US03-26780-1472 Sequence 1472, App
41	83.5	16.2	1833	5	US-09-614-150A-40215 Sequence 40215, App
42	83.5	16.2	1907	1	PCT-US02-24483-64 Sequence 64, App
43	83.5	16.2	2154	1	PCT-US03-26780-1471 Sequence 1471, App
44	83	16.1	171	5	US-09-980-116-88 Sequence 88, App
45	83	16.1	929	1	PCT-US03-30720-2132 Sequence 2132, App

#### ALIGNMENTS

RESULT 1  
US-09-979-932A-606  
Sequence 606, Application US/09979932A  
GENERAL INFORMATION:  
APPLICANT: Odate, Yuichi  
TITLE OF INVENTION: BREAST, GASTRIC, AND PROSTATE CANCER ASSOCIATED ANTIGENS AND  
FILE REFERENCE: L00461.70122, US  
CURRENT APPLICATION NUMBER: US/09/979, 932A  
CURRENT FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 60/136, 526  
PRIOR FILING DATE: 1999-05-28  
PRIOR APPLICATION NUMBER: US 60/153, 454  
PRIOR FILING DATE: 1999-09-10  
NUMBER OF SEQ ID NOS: 1332  
SOFTWARE: Patent version 3.2  
SEQ ID NO 606  
LENGTH: 328  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-979-932A-606

Query Match 100.0%; Score 516; DB 5; Length 328;  
Best Local Similarity 100.0%; Pred. No. 9.2e-32;  
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MANDAVLKRLKQKGAADQIIIEYIKQOVSLKKEKAIQATLRBEKKURVENAKKKEIE 60  
DB 17 MANDAVLKRLKQKGAADQIIIEYIKQOVSLKKEKAIQATLRBEKKURVENAKKKEIE 76  
QY 61 ELKQELIAEILONGVKQAPSGTPIHANSWSENVIOSTAVTVSSG 108  
DB 77 ELKQELIAEILONGVKQAPSGTPIHANSWSENVIOSTAVTVSSG 124

RESULT 2  
PCT-US03-33947-1087  
Sequence 1087, Application PC/TUS0333947  
GENERAL INFORMATION:  
APPLICANT: FIVERINE THERAPEUTICS, INC.  
TITLE OF INVENTION: HUMAN POLYPEPTIDES ENCODED BY POLYNUCLEOTIDES AND METHODS  
FILE REFERENCE: 08940.0017-00304  
CURRENT APPLICATION NUMBER: PCT/US03/33947  
CURRENT FILING DATE: 2003-10-24  
PRIOR APPLICATION NUMBER: US 60/421, 080  
PRIOR FILING DATE: 2002-12-25  
PRIOR APPLICATION NUMBER: US 60/421, 061  
PRIOR FILING DATE: 2002-10-25

PRIOR APPLICATION NUMBER: US 60/421,614  
PRIOR FILING DATE: 2002-10-25  
PRIOR APPLICATION NUMBER: US 60/421,552  
PRIOR FILING DATE: 2002-10-25  
PRIOR APPLICATION NUMBER: US 60/422,178  
PRIOR FILING DATE: 2002-10-30  
PRIOR APPLICATION NUMBER: US 60/422,177  
PRIOR FILING DATE: 2002-10-30  
PRIOR APPLICATION NUMBER: US 60/426,384  
PRIOR FILING DATE: 2002-11-15  
PRIOR APPLICATION NUMBER: US 60/426,394  
PRIOR FILING DATE: 2002-11-15  
PRIOR APPLICATION NUMBER: US 60/426,355  
PRIOR FILING DATE: 2002-11-15  
PRIOR APPLICATION NUMBER: US 60/426,430  
PRIOR FILING DATE: 2002-11-15  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 2352  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 1087  
LENGTH: 123  
TYPE: PRT  
ORGANISM: Homo sapiens  
PCT-US03-33947-1087

Query Match 56.1%; Score 289.5; DB 1; Length 123;  
Best Local Similarity 70.8%; Pred. No. 3.1e-15;  
Matches 66; Conservative 5; Mismatches 16; Indels 7; Gaps 3;

Qy 1 MANNDAVLKRLKQKGAADQIIIEYLKQVSLKEKAILQATL--REKKLRVENAKLKKE 58  
Db 1 MANNDAVLKRLKQKGAADQIIIEYLKQVSLKEKAILQATL--REKKLRVENAKLKKE 56  
Qy 59 IEELKQELIQAETONGVKQIAFPSSGTFPLHANSVSE 94  
Db 57 V-ELKQELIQAETONGVKQIAFPSSGTFPLHANSVSE 91

## RESULT 3

US-09-614-150A-4422  
Sequence 4422, Application US/09614150A  
GENERAL INFORMATION:  
APPLICANT: Venter, J. Craig  
TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID  
TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE  
FILE REFERENCE: CL000728  
CURRENT APPLICATION NUMBER: US/09/614,150A  
PRIOR FILING DATE: 2000-07-11  
PRIOR APPLICATION NUMBER: 60/157,832  
PRIOR FILING DATE: 1999-10-05  
PRIOR APPLICATION NUMBER: 60/160,191  
PRIOR FILING DATE: 1999-10-19  
PRIOR APPLICATION NUMBER: 60/161,932  
PRIOR FILING DATE: 1999-10-28  
PRIOR APPLICATION NUMBER: 60/164,769  
PRIOR FILING DATE: 1999-11-12  
PRIOR APPLICATION NUMBER: 60/173,383  
PRIOR FILING DATE: 1999-12-28  
PRIOR APPLICATION NUMBER: 60/175,693  
PRIOR FILING DATE: 2000-01-12  
PRIOR APPLICATION NUMBER: 60/184,831  
PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: 60/191,637  
PRIOR FILING DATE: 2000-03-23  
NUMBER OF SEQ ID NOS: 43008  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 4422  
LENGTH: 294  
TYPE: PRT  
ORGANISM: DROSOPHILA  
US-09-614-150A-4422

Query Match 20.5%; Score 106; DB 5; Length 294;  
Best Local Similarity 31.6%; Pred. No. 0.33;  
Matches 24; Conservative 21; Mismatches 27; Indels 4; Gaps 1;

Qy 6 AVLKRLKQKGAADQIIIEYLKQVSLKEKAILQATLREKKLRVENAKLKKEIEELKOE 65  
Db 2 ADLQIASNNRERBAIISIGIQ-----QLVERQKQELIKENALAKEVEALAAQ 57  
Qy 66 LQAETONGVKQIAFP 81  
Db 58 LVQLELRNGKQIIVP 73

## RESULT 4

US-09-614-150A-13074  
Sequence 13074, Application US/09614150A  
GENERAL INFORMATION:  
APPLICANT: Venter, J. Craig  
TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID  
TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE  
FILE REFERENCE: CL000728  
CURRENT APPLICATION NUMBER: US/09/614,150A  
PRIOR FILING DATE: 2000-07-11  
PRIOR APPLICATION NUMBER: 60/157,832  
PRIOR FILING DATE: 1999-10-05  
PRIOR APPLICATION NUMBER: 60/160,191  
PRIOR FILING DATE: 1999-10-19  
PRIOR APPLICATION NUMBER: 60/161,932  
PRIOR FILING DATE: 1999-10-28  
PRIOR APPLICATION NUMBER: 60/164,769  
PRIOR FILING DATE: 1999-11-12  
PRIOR APPLICATION NUMBER: 60/173,383  
PRIOR FILING DATE: 1999-12-28  
PRIOR APPLICATION NUMBER: 60/175,693  
PRIOR FILING DATE: 2000-01-12  
PRIOR APPLICATION NUMBER: 60/184,831  
PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: 60/191,637  
PRIOR FILING DATE: 2000-03-23  
NUMBER OF SEQ ID NOS: 43008  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 13074  
LENGTH: 1975  
TYPE: PRT  
ORGANISM: DROSOPHILA  
US-09-614-150A-13074

Query Match 18.6%; Score 96; DB 5; Length 1975;  
Best Local Similarity 26.1%; Pred. No. 10;  
Matches 36; Conservative 15; Mismatches 41; Indels 46; Gaps 4;

Qy 11 LKQKGAADQIIIEYLKQVSLKEKAILQATLREKKLRVENAKLKKEIEELKOE 37  
Db 679 LKQKGAADQIIIEYLKQVSLKEKAILQATLREKKLRVENAKLKKEIEELKOE 738  
Qy 38 LQATLREKKLRVENAKLKKEIEELKOE-----LQAETONGVKQIAFPSSGTFPLHANS 90  
Db 739 LQATLREKKLRVENAKLKKEIEELKOE-----LQAETONGVKQIAFPSSGTFPLHANS 792  
Qy 91 MSENVIQSTAVTVSSG 108  
Db 793 LKQKGAADQIIIEYLKQVSLKEKAILQATLREKKLRVENAKLKKEIEELKOE 810

## RESULT 5

PCT-US03-38193-1546  
Sequence 1546, Application PC/TUS0338193  
GENERAL INFORMATION:  
APPLICANT: Aziz, Natasha  
APPLICANT: Gineburg, Wendy M.

```

; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; FILE REFERENCE: 05882.0193.00PC00
; CURRENT APPLICATION NUMBER: PCT/US03/38193
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1546
; LENGTH: 1046
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US03-38193-1546

```

```

Query Match      17.9%; Score 92.5; DB 1; Length 1046;
Best Local Similarity 26.9%; Pred. No. 10;
Matches 25; Conservative 30; Mismatches 31; Indels 7; Gaps 2;

```

```

Qy      8 LKRLQKGAADQIIEYKQVSLKKEKALQATLREKKLRVENAKLKEIEELKQELI 67
Db      735 LKRLDVEYRGQAQAIIEFLKEQISLAEEKML-----DYERLQRAEAQKQGVESLRKRL 788

```

```

Qy      68 QAEIQ-NGVKQIAFPSTPLHANSVSENVIOS 99
Db      789 VAENRLQAVBALCSSQHTHMIESNDISEETIRT 821

```

```

RESULT 6
US-10-723-860-1546
; Sequence 1546, Application US/10723860
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1546
; LENGTH: 1046
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-723-860-1546

```

```

Query Match      17.9%; Score 92.5; DB 6; Length 1046;
Best Local Similarity 26.9%; Pred. No. 10;
Matches 25; Conservative 30; Mismatches 31; Indels 7; Gaps 2;

```

```

Qy      8 LKRLQKGAADQIIEYKQVSLKKEKALQATLREKKLRVENAKLKEIEELKQELI 67
Db      735 LKRLDVEYRGQAQAIIEFLKEQISLAEEKML-----DYERLQRAEAQKQGVESLRKRL 788

```

```

Qy      68 QAEIQ-NGVKQIAFPSTPLHANSVSENVIOS 99
Db      789 VAENRLQAVBALCSSQHTHMIESNDISEETIRT 821

```

```

RESULT 7
US-10-258-899A-1169
; Sequence 1169, Application US/10258899A
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Dmanac, Radoje T.
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhou, Ping

```

```

; APPLICANT: Xu, Chongjun
; APPLICANT: Cao, Yicheng
; APPLICANT: Ma, Yungting
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Wang, Zhi Wei
; APPLICANT: Xue, Aidong
; APPLICANT: Yang, Yonghong
; APPLICANT: Wehrman, Tom
; APPLICANT: Goodrich, RyLe
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 787CIP2-2G/US
; CURRENT APPLICATION NUMBER: US/10/258,899A
; CURRENT FILING DATE: 2003-11-06
; PRIOR APPLICATION NUMBER: PCT/US01/04098
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 09/774,434
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 09/728,422
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 09/693,325
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 09/663,561
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: 09/654,936
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 09/620,325
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/598,075
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 3960
; SOFTWARE: Cubetom
; SEQ ID NO 1169
; LENGTH: 1047
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-258-899A-1169

```

```

Query Match      17.9%; Score 92.5; DB 6; Length 1047;
Best Local Similarity 26.9%; Pred. No. 10;
Matches 25; Conservative 30; Mismatches 31; Indels 7; Gaps 2;

```

```

Qy      8 LKRLQKGAADQIIEYKQVSLKKEKALQATLREKKLRVENAKLKEIEELKQELI 67
Db      736 LKRLDVEYRGQAQAIIEFLKEQISLAEEKML-----DYERLQRAEAQKQGVESLRKRL 789

```

```

Qy      68 QAEIQ-NGVKQIAFPSTPLHANSVSENVIOS 99
Db      790 VAENRLQAVBALCSSQHTHMIESNDISEETIRT 822

```

```

RESULT 8
US-10-258-899A-3137
; Sequence 3137, Application US/10258899A
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Dmanac, Radoje T.
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhou, Ping
; APPLICANT: Xu, Chongjun
; APPLICANT: Cao, Yicheng
; APPLICANT: Ma, Yungting
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wang, Dunrui

```

```
APPLICANT: Wang, Jian-Rui
APPLICANT: Zhang, Jie
APPLICANT: Ren, Feiyan
APPLICANT: Chen, Rui-hong
APPLICANT: Wang, Zhi Wei
APPLICANT: Xue, Aiqiong
APPLICANT: Yang, Yonghong
APPLICANT: Wehrman, Tom
APPLICANT: Goodrich, Ryle
TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
FILE REFERENCE: 787CIP-26/US
CURRENT APPLICATION NUMBER: US/10/258, 899A
PRIOR FILING DATE: 2003-11-06
PRIOR APPLICATION NUMBER: PCT/US01/04098
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: 09/774,434
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: 09/728,422
PRIOR FILING DATE: 2000-11-30
PRIOR APPLICATION NUMBER: 09/693,325
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 09/663,561
PRIOR FILING DATE: 2000-09-15
PRIOR APPLICATION NUMBER: 09/654,936
PRIOR FILING DATE: 2000-09-01
PRIOR APPLICATION NUMBER: 09/620,325
PRIOR FILING DATE: 2000-07-19
PRIOR APPLICATION NUMBER: 09/598,075
PRIOR FILING DATE: 2000-06-20
PRIOR APPLICATION NUMBER: 09/560,875
PRIOR FILING DATE: 2000-04-27
PRIOR APPLICATION NUMBER: 09/496,914
PRIOR FILING DATE: 2000-02-03
NUMBER OF SEQ ID NOS: 3960
SOFTWARE: Custom
SEQ ID NO 3137
LENGTH: 1047
TYPE: PRT
ORGANISM: Homo sapiens
US-10-258-899A-3137
```

```
Query Match 17.3%; Score 92.5; DB 6; Length 1047;
Best Local Similarity 26.9%; Pred. No. 10;
Matches 25; Conservative 30; Mismatches 31; Indels 7; Gaps 2;

QY 8 LKLEQKGAADQIIEYLKQVSLKEKAILQATLREKKLRVENAKLKEIELEKQELI 67
DB 736 LKLEQKGAADQIIEYLKQVSLKEKAILQATLREKKLRVENAKLKEIELEKQELI 789
QY 68 QAEIQ-NGVQKQIAFPSTGPLHANSMSVENVIOSTAVTTVS 99
DB 790 VAEIRLQAEALCEALSSQHTHMIENSDISEETIRT 822
```

```
RESULT 9
US-10-330-773-712
Sequence 712, Application US/10330773
GENERAL INFORMATION:
APPLICANT: David W. Morris
APPLICANT: Marc Malandro
TITLE OF INVENTION: Novel Compositions and Methods in Cancer
FILE REFERENCE: 529452001300
CURRENT APPLICATION NUMBER: US/10/330,773
CURRENT FILING DATE: 2002-12-27
NUMBER OF SEQ ID NOS: 981
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 712
LENGTH: 911
TYPE: PRT
ORGANISM: Homo sapiens
US-10-330-773-712
Query Match 17.4%; Score 90; DB 6; Length 911;
```

```
Best Local Similarity 26.1%; Pred. No. 14;
Matches 29; Conservative 25; Mismatches 43; Indels 14; Gaps 4;

QY 6 AVLKRLKQKGAAD---QIIEYLKQVSLKEKAILQATLREKK-----LRVENAKL 55
DB 379 SLVRLDQKDEINQOSQLAEKIQO--MLDQDELASTRRDYEKIQBELTRLQIENEA 436
QY 56 KKEIELEKQELIQAEIQNGVQKQIAFPSTGPLHANSMSVENVIOSTAVTTVS 106
DB 437 KQVEKVEIQAELELAVNVDQKSGVEEDKT--RANQQLDELQAKTTTLITTT 485
```

```
RESULT 10
US-10-330-773-709
Sequence 709, Application US/10330773
GENERAL INFORMATION:
APPLICANT: David W. Morris
APPLICANT: Marc Malandro
TITLE OF INVENTION: Novel Compositions and Methods in Cancer
FILE REFERENCE: 529452001300
CURRENT APPLICATION NUMBER: US/10/330,773
CURRENT FILING DATE: 2002-12-27
NUMBER OF SEQ ID NOS: 981
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 709
LENGTH: 952
TYPE: PRT
ORGANISM: Mus musculus
US-10-330-773-709
```

```
Query Match 17.4%; Score 90; DB 6; Length 952;
Best Local Similarity 26.1%; Pred. No. 15;
Matches 29; Conservative 25; Mismatches 43; Indels 14; Gaps 4;

QY 6 AVLKRLKQKGAAD---QIIEYLKQVSLKEKAILQATLREKK-----LRVENAKL 55
DB 420 SLVRLDQKDEINQOSQLAEKIQO--MLDQDELASTRRDYEKIQBELTRLQIENEA 477
QY 56 KKEIELEKQELIQAEIQNGVQKQIAFPSTGPLHANSMSVENVIOSTAVTTVS 106
DB 478 KQVEKVEIQAELELAVNVDQKSGVEEDKT--RANQQLDELQAKTTTLITTT 526
```

```
RESULT 11
US-10-679-063-2630
Sequence 2630, Application US/10679063
GENERAL INFORMATION:
APPLICANT: Egerton, Michael D
TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
FILE REFERENCE: 38-15(52054)B
CURRENT APPLICATION NUMBER: US/10/679,063
CURRENT FILING DATE: 2003-10-02
PRIOR APPLICATION NUMBER: 60/415,758
PRIOR FILING DATE: 2002-10-02
NUMBER OF SEQ ID NOS: 27373
SEQ ID NO 2630
LENGTH: 300
TYPE: PRT
ORGANISM: Glycine max
US-10-679-063-2630
```

```
Query Match 17.3%; Score 89.5; DB 6; Length 300;
Best Local Similarity 26.7%; Pred. No. 5.7;
Matches 28; Conservative 25; Mismatches 43; Indels 9; Gaps 2;

QY 2 ANNDVAVLKLEQKGAADQIIEYLKQVSLKEKAILQATLREKKLRVENAKLKEIELE 61
DB 123 STNPADMKRLRKRVNDR-----SARRSSAREQQLSELQVEKLVKNNAITVKQPTD 176
QY 62 LKQELIQAEIQNGVQKQIAFPSTGPLHANSMSVENVIOSTAVTTVS 106
DB 177 ASQHFREADTNNRVLSKDVEA--LRKVKLAEDMWTSSFTTLN 218
```

```
RESULT 12
US-09-614-150A-31653
; Sequence 31653, Application US/09614150A
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CU00728
; CURRENT APPLICATION NUMBER: US/09/614,150A
; CURRENT FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31653
; LENGTH: 1305
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-09-614-150A-31653

Query Match          16.9%; Score 87; DB 5; Length 1305;
Best Local Similarity 30.2%; Pred. No. 33;
Matches 32; Conservative 22; Mismatches 38; Indels 14; Gaps 4;

Qy      2  ANDAVLK-RLBQKGAEDQIIEYLKQVSLKREKALIQATRE-----EKKLRVEN 52
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      37  SNOVVAKRSRLARIQNDLIRQVQLQANQQLADLQRTAQEVNDLYGKKEQQRRL 96
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Qy      53  AKLKEKEE-----LKQEL-IOAEIQNGVYKQIAFSGTFLHANSWS 93
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      97  EKRTQIGERGQLEKELDVQVGCNENLQEQLOVRGLFVEAKDVLS 142
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

RESULT 13
US-10-258-899A-3150
; Sequence 3150, Application US/10258899A
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Drmanac, Radoje T.
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhou, Ping
; APPLICANT: Xu, Chongjun
; APPLICANT: Cao, Yicheng
; APPLICANT: Ma, Yunguang
; APPLICANT: Wang, Duntui
; APPLICANT: Wang, Qiang A.
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Ruihong
; APPLICANT: Wang, Zhi Wei
; APPLICANT: Xue, Aidong
; APPLICANT: Yang, Yonghong
; APPLICANT: Wehrman, Tom
; APPLICANT: Goodrich, Ryle
```

```
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 787CIP2-2G/US
; CURRENT APPLICATION NUMBER: US/10/258, 899A
; CURRENT FILING DATE: 2003-11-06
; PRIOR APPLICATION NUMBER: PCT/US01/04098
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 09/774,434
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 09/728,422
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 09/693,325
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 09/663,561
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: 09/654,936
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 09/620,325
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/598,075
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 3960
; SOFTWARE: Cuslom
; SEQ ID NO 3150
; LENGTH: 931
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-258-899A-3150

Query Match          16.8%; Score 86.5; DB 6; Length 931;
Best Local Similarity 28.6%; Pred. No. 26;
Matches 28; Conservative 19; Mismatches 28; Indels 23; Gaps 4;

Qy      3  NDAVLRLBQKGAEDQIIEYLKQ--QVS-----LKEKALIQATRE-----KLR 49
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      476  NSERQIILQDENHNLQVYSSLRQSQISAEARVKQIEKENKLHHSIKTSKSLKIE 535
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Qy      50  VNAKLKEE-----EELKQELIOAEIQNGVQ 77
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      536  FEKQIKKELEHYKKEGARAELLENLHLKEKEHLLQ 573
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

RESULT 14
US-10-258-899A-1182
; Sequence 1182, Application US/10258899A
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Drmanac, Radoje T.
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhou, Ping
; APPLICANT: Xu, Chongjun
; APPLICANT: Cao, Yicheng
; APPLICANT: Ma, Yunguang
; APPLICANT: Wang, Duntui
; APPLICANT: Wang, Qiang A.
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Ruihong
; APPLICANT: Wang, Zhi Wei
; APPLICANT: Xue, Aidong
; APPLICANT: Yang, Yonghong
; APPLICANT: Wehrman, Tom
; APPLICANT: Goodrich, Ryle
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 787CIP2-2G/US
; CURRENT APPLICATION NUMBER: US/10/258, 899A
; CURRENT FILING DATE: 2003-11-06
; PRIOR APPLICATION NUMBER: PCT/US01/04098
```

Search completed: January 2, 2004, 18:30:13  
Job time : 13.7059 secs

;; PRIOR FILING DATE: 2001-02-05  
;; PRIOR APPLICATION NUMBER: 09/774,434  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: 09/728,422  
;; PRIOR FILING DATE: 2000-11-30  
;; PRIOR APPLICATION NUMBER: 09/693,325  
;; PRIOR FILING DATE: 2000-10-20  
;; PRIOR APPLICATION NUMBER: 09/663,561  
;; PRIOR FILING DATE: 2000-09-15  
;; PRIOR APPLICATION NUMBER: 09/654,936  
;; PRIOR FILING DATE: 2000-09-01  
;; PRIOR APPLICATION NUMBER: 09/620,325  
;; PRIOR FILING DATE: 2000-07-19  
;; PRIOR APPLICATION NUMBER: 09/598,075  
;; PRIOR FILING DATE: 2000-06-20  
;; PRIOR APPLICATION NUMBER: 09/560,875  
;; PRIOR FILING DATE: 2000-04-27  
;; PRIOR APPLICATION NUMBER: 09/496,914  
;; PRIOR FILING DATE: 2000-02-03  
;; NUMBER OF SEQ ID NOS: 3960  
;; SOFTWARE: Custom  
;; SEQ ID NO: 1182  
;; LENGTH: 990  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-10-258-899A-1182

Query Match 16.8%; Score 86.5; DB 6; Length 990;  
Best Local Similarity 28.6%; Pred. No. 28;  
Matches 28; Conservative 19; Mismatches 28; Indels 23; Gaps 4;

OY 3 NNDVLRLEKRGADQIIEYLKQ--QVS-----LTKKAILQATLRRE---KCLR 49  
DB 535 NSERQIKLEQENHNLQTVSSLRQSQISAEARVKDIEKENKILHESIKETSSKSLKIE 594  
OY 50 VENAKLKEI-----EELKQELIQAEIQNGVKQ 77  
DB 595 FEKQIKKELEHYKEKGERAELEENLEHLEKENELLQ 632

RESULT 15  
US-10-425-114A-68086  
; Sequence 68086, Application US/10425114A  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jindong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovacic, David K.  
; APPLICANT: Screen, Steven E  
; APPLICANT: Tabaska, Jack E  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(5313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114A  
; NUMBER OF SEQ ID NOS: 2003-04-28  
; SEQ ID NO 68086  
; LENGTH: 680  
; TYPE: PRT  
; ORGANISM: Zea mays subsp. mexicana  
; FEATURE:  
; OTHER INFORMATION: Clone ID: UC-ZMROTOSINTE068G03\_FLI.pep  
US-10-425-114A-68086

Query Match 16.7%; Score 86; DB 6; Length 680;  
Best Local Similarity 39.0%; Pred. No. 22;  
Matches 23; Conservative 10; Mismatches 20; Indels 6; Gaps 1;

OY 12 EKGAEADQIIEYLKQGVSLKEKAILQATLRREKCLRVENAKLKEIIEELKQELIQAE 70  
DB 52 ESKSRKAEALLVKNLESVSLKKG-----LEAKYLEDRLEAEKMIIEELKSEIANAQ 104